AD_____

Award Number: W81XWH-09-2-0175

TITLE: Proposal for Development of EBM-CDSS (Evidence-Based Clinical Decision Support System) To Aid Prognostication in Terminally Ill Patients

PRINCIPAL INVESTIGATOR: Benjamin Djulbegovic, M.D., Ph.D.

CONTRACTING ORGANIZATION: University of South Florida
Tampa, FL 33620

REPORT DATE: October 2011

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2011	Annual	25 September 2010 – 24 September 2011

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Proposal for Development of EBM-CDSS (Evidence-Based Clinical Decision Support System) To Aid Prognostication in Terminally Ill Patients	**5b. GRANT NUMBER** W81XWH-09-2-0175
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Benjamin Djulbegovic	**5e. TASK NUMBER**
E-Mail: bdjulbeg@health.usf.edu	**5f. WORK UNIT NUMBER**

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of South Florida Tampa, FL 33620	

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Goal of the project is to develop an Evidence-based Clinical Decision Support (CDSS-EBM) system and make it available at the point of care to improve prognostication of the life expectancy of terminally ill patients to improve referral ofpatients to hospice. We have obtained final approvals from the USAMRMC Office of Research Protections (ORP) Human Research Protections Office (HRPO). So far our key research accomplishments are as follows; we

☐ Extracted data from patient charts of 590 (deceased) patients who were admitted to TampaBay Lifepath Hospice Center.

☐ Developed an electronic database.

☐ Cleaned and entered the data in the database.

☐ Calculated Charlson co-morbidity index for each of the 590 patients. The Charlson co-morbidity index was used to calculate life expectancy estimates according to the DEALE prognostication model [3].

☐ Conducted external validation of SUPPORT, DEALE, PPS, ECOG, Karnofsky prognostication models using the collected data.

☐ Developed decision curves for SUPPORT, DEALE, PPS, ECOG, Karnofsky prognostication models.

☐ Developed the Rough Set Theory prognostication model using the data collected from the TampaBay Lifepath Hospice center.

☐ Refined our web-based version of the CDSS-EBM for hospice referral.

☐ Drafted a manuscript addressing prognosis in lung cancer using a novel data analysis methodology.

☐ Conducted search, critical appraisal and created evidence profiles for the drugs used for pain management.

☐ Refined our evidence based pain management module and incorporated the above-mentioned evidence profiles.

☐ Drafted and submitted 3 manuscripts for peer-reviewed publication and one conference abstract.

15. SUBJECT TERMS
CDSS, SUPPORT, DEALE, Terminally ill, Hospice, Prognostication

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	44	**19b. TELEPHONE NUMBER** *(include area code)*

Table of contents	
	Page number
Introduction	4
Work done so far	4
Key Research Accomplishments	5
Reportable Outcomes	6
Conclusion	6
References	7
Appendices	8

Introduction

The goal of this project is to develop an Evidence-based Clinical Decision Support System (CDSS-EBM) available at the point of care which will improve prognostication of life expectancy of terminally ill patients and facilitate the hospice referral process. In addition, the CDSS-EBM will be expanded with an evidence based pain management module (EB-PMM) to assist physicians managing patients with pain.

Work done so far:

We submitted the required documents including the research protocol and informed consent forms to the regulatory specialist with the Telemedicine and Advanced Technology Research Center (TATRC), and secured the approval from the USAMRMC Office of Research Protections (ORP) Human Research Protections Office (HRPO).

We have removed James A. Haley Veteran's Hospital as one of our study sites. We have revised our study protocol and related study documents such as informed consent forms to reflect this change. We submitted an amendment request to reflect this change in study sites to the University of South Florida's (USF) institutional review board (IRB) and have obtained authorization from USF IRB office.

We completed the first retrospective phase of this project.

Our progress regarding each of the tasks outlined in the statement of work is as follows:

Task 1: We extracted, cleaned and entered data from 590 (deceased) patient charts from the Tampa Bay LifePath Hospice Center into an electronic database. Data extracted included all variables used in the SUPPORT [1], PPS, ECOG, Karnofsky prognostication models.

Task 2: We have completed the calibration and external validation of SUPPORT, DEALE, PPS, ECOG, Karnofsky prognostication models using the data collected from Tampa Bay LifePath Hospice Center. Specifically, the models were tested against observed survival duration. PPS, Karnofsky and ECOG risk scores were predicted using a flexible family of Royston-Parmar parametric models and adjusted for age, gender and presence of cancer. We utilized several metrics to assess the performance of these models. Specifically, we used the Brier score and scaled Brier score (which is very similar to the Pearson correlation coefficient R^2), the area under the receiver operating characteristic curve (AUROC), and the Hosmer-Lemshow goodness-of-fit p-value (HL).
Brier scores were consistently below the non-informative level of 0.25 and AUROC significantly higher than the non-informative level of 0.5 for the adjusted PPS, Karnofsky and ECOG models. The HL p-value was consistently greater than 0.1 only for PPS. SUPPORT and DEALE models did not predict fit our data well for survival at day one and month one, two and six. The AUROC takes a value close to 0.5, even though the Brier scores were relatively low and HL p-value greater than 0.05, this value is significantly close to 0.5 for SUPPORT and DEALE models.
In summary, none of the prognostication models accurately predicated survival among our cohort of terminally ill patients. However, PPS consistently performed best in predicting survival in terminally ill patients followed by Karnofsky and ECOG.

Based on these analyses we have submitted a manuscript titled "a flexible alternative to the Cox proportional hazard model for prognosticating patient survival" to the Journal of Pain and Symptom Management (see appendix 1). We have also submitted our findings to the American Society of Hematology Annual Meeting as an abstract titled "external validation of prognostic models in terminally ill patients" (see appendix 2)

We have also computed the decision curves related to the SUPPORT, PPS, ECOG and Karnofsky model and the alternatives of referring all, or no patients to hospice. The decision curves are used in regret based decision curve analysis (DCA) [2] where each alternative strategy is compared based on the decision maker's personal preferences as they are described by a threshold probability. We found that for this hospice population, the least regretful action and therefore the optimal strategy was indeed for all patients to be referred to hospice.

Finally, we have also used the data collected from the Tampa Bay LifePath Hospice Center to develop an initial version of a Rough Set Theory (RST) [4] prognostication model for life expectancy. This version of the RST model uses the majority of the variables included in the SUPPORT model and it has shown promising prediction results. Further investigation is required to determine the impact of reducing the number of variables utilized or the addition of qualitative variables such as physicians' choices such as continuation of treatment or referral to hospice. This will be possible when prospective phase of the study is commenced.

Task 3: We have developed a pilot web based version of the CDSS-EBM to compute life expectancy for terminally ill patients based on the DEALE and SUPPORT prognostication models. This CDSS-EBM will be utilized to suggest optimal strategies for particular simulated patients. This pilot CDSS-EBM will facilitate the hospice referral procedure in three steps. First, the probabilities of survival and death for a specific patient are computed based on the SUPPORT prognostication model. Then, based on interviews with either the patient, his/her family, or his/her physician the system elicits personal preferences regarding continuation of life-sustaining vs. palliative care. Finally, using regret DCA, the optimal decision for the specific patient is suggested.

Task 4: We conducted a systematic review addressing prognosis of patients with metastatic lung cancer. We are in the final steps of putting together this manuscript for publication in a peer reviewed journal.

Task 5: We have refined our Evidence-based Chronic Pain Management Module to complement the CDSS-EBM. Our objective is to develop a reliable dosage conversion system as well as a knowledge based for each available pain medication. We have also incorporated evidence profiles for each drug to support the decision making using our pain management module. We have also created a survey to test usefulness of EB-PMM its users. The system is currently going through the final programming phase and will be tested internally and in the clinic in the prospective phase of the study. We have also created the users' manual for the EB-PMM.

Key research accomplishments

- Extracted, cleaned and entered data from patient charts of 590 (deceased) patients who were admitted to TampaBay Lifepath Hospice Center into an electronic database.

- Conducted external validation of SUPPORT, DEALE, PPS, ECOG, Karnofsky prognostication models using the collected data.

- Developed decision curves for SUPPORT, DEALE, PPS, ECOG, Karnofsky prognostication models.

- Developed the Rough Set Theory prognostication model using the data collected from the TampaBay Lifepath Hospice center.

- Conducted search, critical appraisal and created evidence profiles for the drugs as a part of the EB-PMM development. Drafted the user's manual for the EB-PMM.

- Drafted and submitted 3 manuscripts for peer-reviewed publication and one conference abstract.

Reportable outcomes

1. Journal publications: (appendix 1)
    - E. Gil-Herrera, A. Yalcin, A. Tsalatsanis, L. Barnes, B. Djulbegovic, "Rough Set Theory based Prognostication of Life Expectancy for Terminally Ill Patients," to appear in Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society, 2011.
    - Tsalatsanis, L. Barnes, I. Hozo, B. Djulbegovic, "Extensions to Regret-based Decision Curve Analysis: An application to hospice referral for terminal patients", submitted to BMC medical Informatics and Decision Making, 2011
    - A flexible alternative to the Cox proportional hazard model for prognosticating patient survival Branko Miladinovic, PhD ,Rahul Mhaskar, PhD, Sehwan Kim, PhD, Ronald Schonwetter, MD, Benjamin Djulbegovic, MD, PhD submitted to the Journal of Pain and Symptom Management, 2011

2. Abstract presentation: (appendix 2)
    - External validation of Prognostic Models in Terminally Ill Patients Rahul Mhaskar, Branko Miladinovic, Athanasios Tsalatsanis, Alfred Mbah, Ambuj Kumar, Sehwan Kim, Ronald Schonwetter, and Benjamin Djulbegovic. Submitted to the 2011 American Society of Hematology Annual Meeting

Conclusion

We have already completed the majority of tasks described in the statement of work. We believe that we have closely followed the grant's timeline where we could control the work process. At this point, we are working on fine-tuning the standard operating procedures and finalizing the prognostication model to be used for the prospective phase of the study. Our key research findings so far can be summarized as follows:

- Based on results of external validation of prognostic models so far, PPS appears to be the best fitting model for our patient population.
- Based on the results of the regret DCA methodology for the hospice population, the least regretful action corresponds to sending all patients to hospice.
- Based on the area under the curve measure, the RST model performs better than SUPPORT in patient classification. However, further investigation is required to determine the impact of reducing the number of variables utilized or the addition of qualitative variables such as physicians' values regarding hospice referral.

In the near future we intend to investigate further details regarding the development of the RST prognostication methodology and its application to both SUPPORT and hospice datasets. Our goal is develop the appropriate theoretical framework that will facilitate the hospice referral process based on outcomes of multiple prognostication models. Our plan is to develop an evidence-based decision-support system that will help both better referral to hospice as well help with pain management. Ultimately, the usefulness of our system will depend how well it performs when tested in clinical setting.

Next Steps

Our immediate next step is to finalize the prognostication model to be used for the prospective phase of the study. We are also putting together the manual of standard operating procedures to be followed by our team in the prospective phase of the study. We will enroll patients from the Tampa General Hospital for our prospective phase of the study. We will also pilot test our EB-PMM with physicians at Tampa General Hospital.

References

 [1] The-Support-Investigators. A Controlled Trial to Improve Care for Seriously III Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT). JAMA 1995, 247(20):1591-1598.

[2] Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Med Decis Making. November 1, 2006 2006;26(6):565-574.

[3] Beck JR, Kassirer JP, Pauker SG. A convenient approximation of life expectancy (The "D.E.A.L.E").I. Validation of the method. Am J Med. 1982;73:883-888.

[4] Lin TJ, Cercone N. Rough Sets and Data Mining, Analysis of Imprecise Data: Kluwer Academic Publishers; 1997.

Appendix 1

Rough Set Theory based Prognostication of Life Expectancy for Terminally Ill Patients

Eleazar Gil-Herrera, Ali Yalcin, Athanasios Tsalatsanis, Laura E. Barnes and Benjamin Djulbegovic

Abstract—**We present a novel knowledge discovery methodology that relies on Rough Set Theory to predict the life expectancy of terminally ill patients in an effort to improve the hospice referral process. Life expectancy prognostication is particularly valuable for terminally ill patients since it enables them and their families to initiate end-of-life discussions and choose the most desired management strategy for the remainder of their lives. We utilize retrospective data from 9105 patients to demonstrate the design and implementation details of a series of classifiers developed to identify potential hospice candidates. Preliminary results confirm the efficacy of the proposed methodology. We envision our work as a part of a comprehensive decision support system designed to assist terminally ill patients in making end-of-life care decisions.**

INTRODUCTION

ACCORDING to Medicare regulations, a patient should be referred to hospice if his/her life expectancy is less than 6 months [1]. However, despite the well-documented advantages of hospice services, terminally ill patients do not reap the maximum benefits of hospice care with the majority of them being referred to hospice either prematurely or too late. In general, premature hospice referral is translated to patients losing the opportunity to receive potentially effective treatment, which may have prolonged their lives. Conversely, late hospice referral reduces the quality of life for patients and their families. It is apparent that accurate prognostication of life expectancy is of vital importance for all parties involved in the hospice referral process (e.g. patients, their families, and their physicians).

Here, we propose a novel knowledge discovery methodology developed to identify terminally ill patients with life expectancy less than 6 months. The core of the proposed methodology is Rough Set Theory [2]. The rest of this paper describes implementation details, reports results, and discusses limitations and future directions of our work.

Methodology

Literature Review

Approaches for developing prognostic models for estimating survival for seriously ill patients range from the use of traditional statistical and probabilistic techniques [3]-[6], to models based on artificial intelligence techniques such as neural networks, decision trees and rough set methods [7]-[11]. A recent systematic review of prognostic tools for estimating survival in palliative care highlighted the lack of accurate end-of-life prognostic models [13].

Both statistics based techniques and *AI* based models rely on data that are precisely well defined. However, medical information, which represents patients records that include symptoms and clinical signs, is not always well defined and, therefore, the data are represented with vagueness [14]. Particularly, for this kind of information, it becomes very difficult to classify borderline cases in which very small differences in the value of a variable of interest may completely change categorization and therefore the following decisions can changes dramatically [15]. Moreover, the dataset is presented with inconsistencies in the sense that it is possible to have

E. Gil-Herrera and A. Yalcin are with the Department of Industrial and Management System Engineering, University of South Florida, Tampa, FL 33620, USA(e-mail: eleazar@mail.usf.edu, ayalcin@eng.usf.edu).

A. Tsalatsanis, L. E. Barnes and B. Djulbegovic are with the Center for Evidence Based Medicine and Health Outcomes Research, University of South Florida, Tampa, FL 33612, USA (e-mail: atsalats@health.usf.edu, lbarnes@health.usf.edu, bdjulbeg@health.usf.edu).

TABLE I
CONDITION ATTRIBUTES

Name	Description
meanbp	Mean arterial blood pressure Day 3
wblc	White blood cell count Day 3
hrt	Heart rate Day 3
resp	Respiratory rate Day 3
temp	Temperature (Celsius)
alb	Serum Albumin
bili	Bilirubin
crea	Serum Creatinine
sod	Sodium
pafi	Pa02 / (.01 * FiO2)
ca	Presence of cancer
age	Patient's age
hday	Days in hospital at study admit
dzgroup	Diagnosis group
scoma	SUPPORT coma score based on Glasgow coma scale

more than one patient with the same description but showing different outcomes.

In this work we propose the use of Rough Set Theory (RST) [2] to deal with vagueness and inconsistency in the representation of the dataset. RST provides a mathematical tool for representing and reasoning about vagueness and inconsistency. Its fundamentals are based on the construction of similarity relations between dataset objects from which approximate yet useful solutions are provided.

RST has been used in a number of applications dealing with modeling medical prognosis [9]–[12]. For example, Tsumoto et al. [11], provides a framework to model medical diagnosis rules showing theoretically that the characteristics of medical reasoning reflect the concepts of approximation established in Rough Set Theory. Komorowski et al. [12], show that RST is useful to extract medical diagnosis rules to identify a group of patients for whom performing a test that is costly or invasive is redundant or superfluous in the prognosis of a particular medical condition.

In this paper we describe a RST based knowledge discovery methodology to provide a classifier that properly discriminates patients into two groups, those who survive at least 180 days after evaluation for hospice referral and those who do not. ROSETTA [16] software is used to perform the analysis described in the remainder of the paper.

Dataset

The dataset used in this study consists of the 9105 cases from the SUPPORT (Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments) prognostic model dataset [17]. We consider all variables used in the SUPPORT prognostic model as condition attributes, i.e. the physiologic variables along with the diagnosis groups, age, number of days in the hospital before entering the study, presence of cancer, and neurologic function. Attributes' names and descriptions are listed in Table I.

As the decision attribute in the decision table, we define a binary variable (Yes/No) "deceesed_in_6months" using the following two attributes from the SUPPORT dataset:

- *"death"* which represents the event of death at any time up to NDI date (National Death Index date: Dec 31, 1994).
- *"D.time":* number of days of follow up

The values of the decision attribute are calculated converting the *"D.time"* value in months and comparing against the attribute *"death"* as follows:

- If "D.time" < 6 months and "death" is equal to 1 (the patient died within 6 months) then "deceased_in_6months" is equal to "Yes"
- If "D.time" > 6 months and "death" is equal to 1 (the patient died after 6 months) then "deceased_in_6months" is equal to "No"
- If "D.time" > 6 months and "death" is equal to 0 (the patient did not died after 6 months) then "deceased_in_6months" is equal to "No"

Rough Set Theory

Based on RST, we can formally define the prognostication problem as:

$$T = (U, A \cup \{d\}) \tag{1}$$

where *T* represents the dataset in the form of a table. Each row represents an object and each column represents an attribute. *U* is a non-empty finite set of objects and the set *A* represents a non-empty finite set of attributes called the condition attributes. In our case, an object designates a terminally ill patient and an attribute designates each of the fifteen condition attributes that describe a patient (Table I). Also, for every attribute $a \in A$, the function $a: U \rightarrow V_a$ makes a correspondence between an object in *U* to an attribute value V_a which is called the value set of *a*.

The set *T* incorporates an additional attribute *{d}* called the decision attribute. The system represented by this scheme is called a *decision system*.

Rough Set Theory Based Knowledge Discovery Process

RST based knowledge discovery process requires sequential and parallel use of various mathematical, statistical and soft computing methodologies with the objective of identifying meaningful relationships between condition and decision attributes.

The selection of specific methodologies for knowledge discovery is largely dependent on the considered dataset. We have taken the following steps in our approach:

1) Data preprocessing: If the selected table contains "holes" in the form of missing values or empty cell entries; the table may be processed in various ways to yield a completed table in which all entries are present. The data completion process for SUPPORT data

in [17] is adopted in this work. This enables a meaningful comparison between the two approaches in terms of accuracy. After the preprocessing phase, the number of patients with missing information is reduced by 2 cases. Therefore, and there are 9103 complete cases.

The next step in preprocessing is the discretization process. 13 out of 15 of the conditional attributes are continuous; therefore we transformed them into categorical variables. The discretization process is based on the searching of cuts that determine intervals. This process enables the classifier in obtaining a higher quality of classification rules. We found that using cut-off defined by medical experts is the best alternative for the discretization process. We consider the APACHE III Scoring System [5] for determining the cut-off for the physiologic variables along with the age variable. The remaining variables, not defined in [5] are discretized using Boolean Reasoning Algorithm [18] implemented in the ROSETTA software.

Finally, the dataset is divided into training and testing sets containing 500 and 8605 cases, respectively. The training set is used in the discretization process to obtain the cut-off for the numerical attributes.

2) Reduct Generation: This step reduces the dimensionality of the dataset by removing redundant information and consequently decreases the complexity of the mining process. A reduct is the minimal set of attributes that enable the same classification as the complete set of attributes without loss of information. There are many algorithms for computing reducts. As it will be shown later in this paper, the effect of the reduct generation algorithm to the classification performance is critical. Since the computational complexity of the reduct generation problem is NP-hard [18], various suboptimal techniques have been proposed. The technique most appropriate to the problem at hand is the one that generates the best classification accuracy in the testing dataset. In this work, two approaches are used for reduct generation, both using an algorithm based on dynamic reducts [19] and [20], but considering a special treatment of the decision attribute for the second approach which is explained in the next section.

2.1) Using Dynamic Reducts

Dynamic reducts aim at obtaining the most stable sets of reducts for a given dataset by sampling within this dataset. For example, in an iterative manner different samples of the testing set are selected for which reducts are computed using a genetic algorithm [21] and [22]. The reducts that appear more frequently in these samples are selected as the most stable.

Based on the principle of the dynamic reducts technique, we have randomly selected 100 subdivisions of the training set to use for reduct generation. The actual number of patient profiles included in each subdivision of the training set varies between 50% and 90% of the training dataset.

The reducts for each subdivision as well as the reduct from the complete training set are computed. After applying dynamic reducts to the training set, 229 reducts were obtained from which the set of decision rules are generated.

2.2) Using the decision attribute as condition attribute

We consider the alternative of using the decision attribute *d* as one of the condition attributes and calculating the reducts based on this scheme, hence, the condition attribute set is now expressed by:

$$A = (A \cup \{d\}) \tag{2}$$

The decision attribute used as a condition is intended to represent the prognosis estimate of life expectancy, made by a physician, expressed in terms of the decision classes defined for this problem. Survival prognosis models that incorporate physician estimates are shown to improve both predictive accuracy and the ability to identify patients with high probabilities of survival or death [4]. Again using the dynamic reducts approach, 549 reducts were obtained. The next step is the induction of decision rules.

3) Rule Induction. The ultimate goal of the RST based knowledge discovery methodology is to generate decision rules, which will be used in classifying each patient as surviving or not surviving within the defined period of time. A decision rule has the form: *if A then B (A → B),* where *A* is called the condition and *B* the decision of the rule. Decision rules can be thought of as a formal language for drawing conclusions from data.

The decision rules were generated based on the two aforementioned sets of reducts. After the process of reduct generation, the decision table is presented in a compact shape from which the decision rules are generated, therefore, some decision rules may contain fewer attributes than originally defined for the decision table.

4) Classification. Based on the set of rules generated, we can classify patients as surviving or not surviving the six-month period. However, not all rules are conclusive. Patients with profiles identical to the conditions of the rules are not decisively classified. In addition, there are situations of contradictory rules, e.g. one or more rules classify a patient as surviving and some other rules classify the same patient as dying. To overcome these problems a *standard voting* algorithm [18] is used which allows all rules to participate in the decision process and classify a patient based on majority voting.

Results

This section compares the performance of the classification processes based on the decision rules generated. At this stage of the knowledge discovery methodology, the patients in the training dataset are classified as

survive, not survive or *undefined* based on the induced rules and the classification process described. The results are presented in a confusion matrix form.

The accuracy of each classification model is reported in terms of Area under the Receiver Operating Characteristic curve (AUC). The best possible classification is achieved when AUC is equal to 1, while no classification ability exists when AUC is equal to 0.5.

Tables 2 and 3 present the confusion matrix for the classification model based on reducts generated with and without using the decision attribute as a condition attribute.

The dynamic reducts approach without using the decision attribute as a condition attribute does not show any significant discrimination ability. However, it demonstrates a fairly high level of coverage, being able to classify around 85% of the test cases. As shown in Table 3, the classification performance in terms of AUC when using the decision attribute as a part of the condition attributes is approximately 0.90. Both the specificity and sensitivity scores are tremendously improved. However, the classification coverage in this case is reduced to 70%.

TABLE 2
CONFUSION MATRIX. DYNAMIC REDUCTS. THE CLASSIFIER PRESENTS AUC EQUAL TO 0.55 INDICATING CLASSIFICATION INABILITY

.

		Predicted		
		Not survive	Survive	Undefined
Actual	Not survive	1395	1953	677
	Survive	1410	2542	626

Sensitivity = 0.64
Specificity = 0.42
AUC = 0.55

TABLE 3
CONFUSION MATRIX. IMMD APPROACH. THE CLASSIFIER PRESENTS AUC EQUAL TO 0.90 INDICATING GOOD CLASSIFICATION INABILITY

.

		Predicted		
		Not survive	Survive	Undefined
Actual	Not survive	1999	471	1555
	Survive	312	3245	1021

Sensitivity = 0.91
Specificity = 0.81
AUC = 0.90

Conclusions and Future Work

The SUPPORT model is the "gold standard" model for prognostication of terminally ill patients. The AUC for prediction of survival for 180 days in the SUPPORT study is 0.79, and 0.82 when SUPPORT model is combined with physician's estimates [4].

This initial exercise in applying knowledge discovery methodologies based on rough set theory shows promise in developing a reliable methodology to predict life expectancy. The baseline model using dynamic reducts presents several opprotunities for improvement:

1. Due to the limitations of the ROSETTA software, the size of the training set was limited to 500. The size of the training set is fairly limited considering the number of categories associated with each attribute.

2. One area that needs to be explored is the appropriate weighting of the condition attributes in terms of their impact on the decision variable. The baseline case assumes that all physiological attributes are weighed equally. We believe that a careful weighting of the attributes by consulting an expert will greatly improve the classification accuracy of the approach.

Including the physician's estimate in the prognostication process is an important componenet of our future work. The second classifier which uses the decision attribute as a condition attribute is intended to incorporate the professional opinion of the physician. This classifier performed much better than the baseline model and its accuracy exceeded that of the SUPPORT model. However we note that, in this approach only 70% of the test cases could be classified and more research is required to minimize the number of undefined cases. Furthermore, our model used the decision attribute from a retrospective study for which the decision was known with 100% accuracy. A true test would be to apply this approach to a prospective dataset.

Finally, it is important to remember that regardless of the accuracy of any classifier, medical decisions must take into account the individual patient preferences towards alternative forms of treatments[23]. Therefore, our intent is to incorporate our methodology into a patient-centric decision support system to facilatate the hospice referral process.

REFERENCES

[1] L. R. Aiken, "Dying, Death, and Bereavement," *Allyn and Bacon*, 1985, p. 214.

[2] Z. Pawlak, "Rough Sets: Theoretical Aspects of Reasoning about Data," *Kluwer Academic Publishers*, Norwell, MA. 1992.

[3] D. W. Hosmer Jr., S. Lemeshow, "Applied Survival Analysis: Regression Modeling of Time to Event Data," *John Wiley & Sons*, Chichester, 1999.

[4] W. A. Knaus, F. E. Harrell Jr, J. Lynn, L. Goldman, R. S. Phillips, A. F. Connors Jr, et al, "The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults," *Ann Intern Med.* 1995, pp. 191-203.

[5] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P.G. Bastos, C.A Sirio, D.J Murphy, T. Lotring, A. Damiano, "The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults," *Chest*, vol. 100, no. 6, 1991, pp. 1619-1636.

[6] J. R. Bech, S. G. Pauker, J. E. Gottlieb, K. Klein, J. P. Kassirer, "A convenient approximation of life expectancy (The "D.E.A.LE")," Use in medical decision-making, *Am J Med*. 1982, pp. 889-97.

[7] K. J. Cios, J. Kacprzyc, "Medical Data Mining and Knowledge Discovery," *Studies in Fuzziness and Soft Computing 60,* Physica Verlag, Heidelberg, 2001.

[8] J. F. Lucas-Peter, A. Abu-Hanna, "Prognostic methods in medicine," *Artificial Intelligence in Medicine,* vol. 15, no. 2, Feb. 1999, pp. 105-119.

[9] J. Bazan, A. Osmolski, A. Skowron, D. Slezak, M. Sacauka and J. Wroblewski. "Rough Set Approach to the survival Analysis," *Proceedings of the Third International Conference on Rough Sets and Current Trends in Computing series* , 2002, pp. 522-529.

[10] J. P. Grzymala- Busse, J. W. Grzymala-Busse, Z. S. Hippe, "Prediction of melanoma using rule induction based on rough sets," In: *Proc of SCI'01*, 2001, vol. 7, pp. 523-527.

[11] S. Tsumoto, "Modelling Medical Diagnostic Rules Based on Rough Sets," in *Proceedings of the First International Conference on Rough Sets and Current Trends in Computing (RSCTC '98),* Lech Polkowski and Andrzej Skowron (Eds.). Springer-Verlag, London, UK, 1998, pp. 475-482.

[12] J. Komorowski and A. Øhrn, "Modeling prognostic power of cardiac tests using rough sets," *Artificial intelligence in medicine*, Feb. 1999, vol. 15, no. 2, pp. 167-191.

[13] F. Lau, D. Cloutier-Fisher, C. Kuziemsky, et al. "A systematic review of prognostic tools for estimating survival time in palliative care," *Journal of Palliative Care*, 2007, vol. 23, no. 2, pp. 93-112.

[14] T. Williamson, "Vagueness," London, Routledge, 1994.

[15] B. Djulbegovic, "Medical diagnosis and philosophy of vagueness – uncertainty due to borderline cases," *Ann Intern Med*. 2008.

[16] A. Øhrn, J. Komorowski, "ROSETTA: A Rough Set Toolkit for Analysis of Data," *Proc. Third International Joint Conference on Information Sciences, Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97),* Durham, NC, USA, 1997, March 1-5, vol. 3, pp. 403-407.

[17] Support Datasets Archived At ICPSR

[18] J. G. Bazan, H. S. Nguyen, P. Synak, J. Wroblewski, "Rough set algorithms in classification problem," In: L. Polkowski, S. Tsumoto, T.Y Lin, (Eds.), "Rough set methods and applications: new developments in knowledge discovery in information systems. Studies in Fuzziness and Soft Computing," *Physica-Verlag*, Heidelberg, Germany, 2000, pp. 49-88.

[19] J. Bazan, A. Skowron, P. Synak, "Dynamic reducts as a tool for extracting laws from decision tables," *Proceedings of the Eighth International Symposium on Methodologies for Intelligent Systems. Lecture Notes in Artificial Intelligence 869*, Berlin, Springer-Verlag, 1994, pp. 346-355.

[20] J. Bazan, "Dynamic Reducts and Statistical inference," In *Sixth International conference, Information Procesing and Management of Uncertainty in Knowledge-Based Systems*, Granada, Spain, Universidad de Granda, 1996.

[21] J. Wroblewski, "Finding minimal reducts using genetic algorithms," In *Proc. Second International Joint Conference on Information Sciences*, 1995, pp. 186–189.

[22] D. E. Goldberg, "GA in search, optimization, and machine learning," *Addison-Wesley,* 1989.

[23] A.Tsalatsanis, I. Hozo, A. Vickers, B. Djulbegovic, "A regret theory approach to decision curve analysis: A novel method for eliciting decision maker's preferences and decision making", BMC Medical Informatics and Decision making, vol. 10, issue 51, 2010.

Extensions to Regret-based Decision Curve Analysis: An application to hospice referral for terminal patients

Athanasios Tsalatsanis[1][§], Laura Barnes[1], Iztok Hozo[2], Benjamin Djulbegovic[1,3]

[1]Center for Evidence-based Medicine and Health Outcomes Research, University of South Florida, Tampa, FL
[2]Department of Mathematics, Indiana University Northwest, Gary, IN
[3]H. Lee Moffitt Cancer Center & Research Institute

[§]Corresponding author

Abstract

Background

Despite the well documented advantages of hospice care, most terminally ill patients do not reap the maximum benefit from hospice services, with the majority of them receiving hospice care either prematurely or delayed. Decision systems to improve the hospice referral process are sorely needed.

Methods

We present a novel theoretical framework that is based on well-established methodologies of prognostication and decision analysis to assist with the hospice referral process for terminally ill patients. We linked the SUPPORT statistical model, widely regarded as one of the most accurate models for prognostication of terminally ill patients, with the recently developed *regret based decision curve analysis* (*regret DCA*). We extend the *regret DCA* methodology to consider harms associated with the prognostication test as well as harms and effects of the management strategies. In order to enable patients and physicians in making these complex decisions in real-time, we developed an easily accessible web-based decision support system available at the point of care.

Results

The web-based decision support system facilitates the hospice referral process in three steps. First, the patient or surrogate is interviewed to elicit his/her personal preferences regarding the continuation of life-sustaining treatment vs. palliative care. Then, *regret DCA* is employed to identify the best strategy for the particular patient in terms of threshold probability at which he/she is indifferent between continuation of treatment and of hospice referral. Finally, if necessary, the probabilities of survival and death for the particular patient are computed based on the SUPPORT prognostication model and contrasted with the patient's threshold probability. The web-based design of the CDSS enables patients, physicians, and family members to participate in the decision process from anywhere internet access is available.

Conclusions

We present a theoretical framework to facilitate hospice referral process. Further rigorous clinical evaluation including testing in a prospective randomized controlled trial is required and planned.

Background

Introduction

Hospice services have been proven to provide better quality of care to dying patients[1-3] by optimizing pain relief [4, 5] and reducing emotional stress [1, 6, 7]. Furthermore, hospice care is associated with greater patient-family satisfaction[8], is shown to be cost effective[9, 10], and most importantly, it has been attributed with increased survival in some patients [11]. Despite these well documented advantages, many terminally ill patients do not reap maximum benefits from hospice care. The fundamental reason for this is related to the **less than optimal and frequently poorly timed referral** of terminally ill patients to hospice [1, 12]. As a result, many patients die within a few days of referral, or live many years after the referral was made [13].

According to Medicare regulations, a person should be referred to hospice if his/her "life expectancy (LE) is 6 months or less" [1, 14]. Hence, the problem of meaningful referrals relates to the accurate estimation (prognosis) of death within approximately 6 months after evaluation for hospice care. However, statistical models designed to assist physicians in predicting life expectancy (LE), although beneficial [15, 16], so far they failed to improve the quality of care at the end of life [17-21].

One such statistical model is SUPPORT (Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments), designed to calculate the probability of survival over a period of 180 days [22, 23]. Although the SUPPORT model has been well validated [17, 22] for prognostication of LE in terminally ill patients, a controlled trial of SUPPORT failed to demonstrate any impact on the overall quality of care for these patients [17, 20]. We postulate that this lack of impact may be due to the fact that SUPPORT results, were not linked to any decision methodology that would translate the probability of survival to a hospice referral recommendation. Therefore, the full potential of the model's prognostication power remained unexploited.

In this work, we link the SUPPORT prognostication model with the recently developed decision methodology *regret DCA* [24] to facilitate the hospice referral process. *Regret DCA* relies on regret theory and decision curve analysis [25] to recommend the optimal management strategy for a patient, accounting for the personal attitudes and values of the particular patient or his/her surrogate.

Furthermore, we extend *regret DCA* to incorporate harms and effects of treatment as well as harms associated with the prognostication test to the decision model. The presented methodology is integrated into a comprehensive clinical decision support system developed to facilitate the hospice referral process.

Methods

Dataset

In our analysis, we utilized the entire SUPPORT dataset, both development and validation cohorts. The dataset is presented in detail elsewhere [22]. Medical records of 8,329 seriously ill hospitalized adults are included.

SUPPORT MODEL

SUPPORT is a multivariable model designed to estimate probability of survival for seriously ill hospitalized patients over a period of the subsequent 180 days. The model variables include the patient's medical condition compatible with one of eight major diagnostic groupings (Acute Respiratory Failure, Multiple Organ System Failure, Chronic Obstructive Pulmonary Disease, Congestive Heart Failure, Hepatic Cirrhosis, Neurological Coma, Lung or Colon Cancer), the patient's current age, number of days in the hospital before study entry, neurologic status, and 11 physiologic measures recorded on day 3 after study entry [22].

The SUPPORT implementation for the estimation of survival probability is detailed in the appendix. Due to the nature of the hospice referral problem we also express the survival probability in terms of mortality. We can convert the estimated survival probability (SP) (equation A2) to probability of death within 180 days (denoted here as p) using the equation:

$$p = 1 - SP = 1 - P\{T \geq t \mid \text{disease group} = i\} \qquad (1)$$

where SP is the survival probability computed by SUPPORT, $i \in [1,8]$ the patient's disease group, T is the survival time in days, and t is an arbitrary time (typically expressed in days e.g. $t \in [1,180]$).

In terms of accuracy, the SUPPORT model has an area under the receiver-operating characteristics curve (ROC) for prediction of surviving 180 days of 0.79 in the phase I development cohort and 0.78 in the phase II validation cohort [22].

Decision model

Figure 1 depicts the decision tree summarizing the process of hospice referral. The four outcomes and their corresponding utilities (U) shown are:

1. U_1: Refer the patient to hospice and the patient's LE is less than or equal to 6 months ($Hosp|LE \leq 6$).

2. U_2: Refer the patient to hospice and the patient's LE is greater than 6 months ($Hosp|LE > 6$).

3. U_3: Continue treating the patient and the patient's LE is less than or equal to 6 months ($Rx|LE \leq 6$).

4. U_4: Continue treating the patient and the patient's LE is greater than 6 months ($Rx|LE > 6$).

p is the probability associated with the presence of an event (e.g. patient's LE ≤ 6 months) as predicted by the SUPPORT model, $1 - p$ is the probability associated with the absence of the same event (e.g. patient's LE > 6 months).
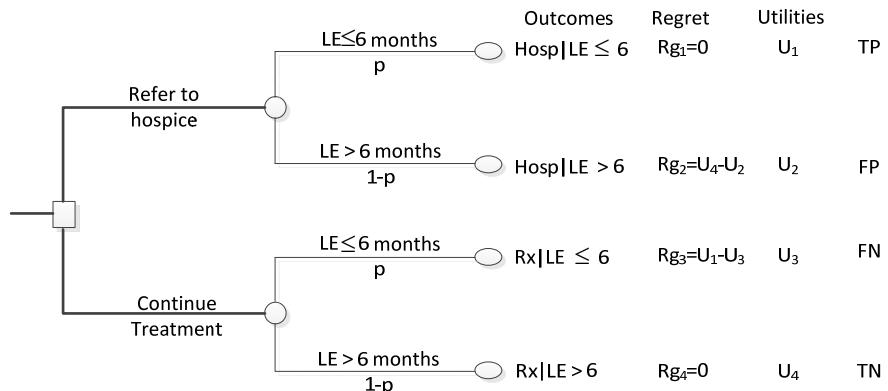

Figure 1. Decision tree for hospice referral. p is the probability that a patient's LE is less than or equal to 6 months; 1-p is the probability that a patient's LE is greater than 6 months; U_i: are the utilities associated with each outcome; Rg_i is the regret associated with each outcome.

As with any decision, one may come to realize that, in retrospect, an alternative decision would have been preferable. This knowledge may bring a sense of loss or regret [26-32]. In this paper, we use this sense of regret to determine the preferences of the decision maker towards alternative management strategies. Specifically, we employ regret theory to estimate the threshold probability, P_t, at which the decision maker (patient, physician, or family member) is indifferent between continuation of treatment vs. hospice referral. Based on the concept of threshold probability, the patient should be referred to hospice if his/her probability of death is greater than or equal to P_t (e.g. $p \geq P_t$), and he/she should continue receiving curative treatment otherwise ($p < P_t$).

The threshold probability is derived as [24]:

$$P_t = \frac{1}{1 + \frac{U_1 - U_3}{U_4 - U_2}} \tag{2}$$

In (2) U_1-U_3 is associated with regret of omission (e.g. the patient was not referred to hospice, instead he/she continued receiving unnecessary treatment) and U_4-U_2 with regret of commission (e.g. the patient was unnecessary referred to hospice instead of continue receiving life-sustaining treatment) [24].

16

To elicit the decision maker's regret, and therefore threshold probability, we utilize the DVAS (Dual Visual Analogue Scale) method [24]. One visual analogue scale is used to capture the regret associated with failing to refer the patient to hospice (e.g continue unnecessary treatment) and the second scale to measure the regret associated with unnecessary hospice referral (e.g. failing to provide life-sustaining treatment) (figure 3).

Elicitation of threshold probability can be achieved through a set of questions such as:

1. *On the scale 0 to 100, where 0 indicates no regret and 100 indicates the maximum regret you could feel, how would you weigh the level of your regret if you were not referred to hospice but instead you continued receiving unnecessary treatment? That is, how much would you regret if you did not reap the benefits of hospice care?* *Note that this value corresponds to $U_1 - U_3$.
2. *On the scale 0 to 100, where 0 indicates no regret and 100 indicates the maximum regret you could feel, how would you weigh the level of your regret, if you were referred to hospice instead of continue receiving necessary life-sustaining treatment? That is, how much would you regret if you sustained harms from hospice care?* *Note that this value corresponds to $U_4 - U_2$.

For example, suppose that the patient – who is aware of his/her terminal condition– answers 50 and 25 to the questions 1 and 2 respectively. This means that the patient considers 50/25=2 times worse not to be referred to hospice when necessary than receiving an unnecessary hospice referral. The threshold probability for this patient is (equation 2)

$$P_t = \frac{1}{1+\frac{U_1-U_3}{U_4-U_2}} = \frac{1}{1+\frac{50}{25}} = 0.33 \text{ or } 33\%.$$

On a scale 0 to 100, where 0 indicates no regret and 100 indicates the maximum regret you could feel, how would you rate the level of your regret if you were not referred to hospice but instead you continued receiving unnecessary treatment?

No regret 0 ←————————————50————————————→ 100 Maximum regret

This value represents the consequences of failure to refer to hospice

On a scale 0 to 100, where 0 indicates no regret and 100 indicates the maximum regret you could feel, how would you rate the level of your regret if you were referred to hospice instead of continue receiving necessary life-sustaining treatment?

No regret 0 ←————25————————————————————→ 100 Maximum regret

This value represents the consequences of unnecessary hospice referral
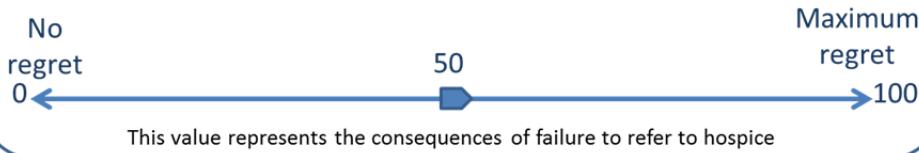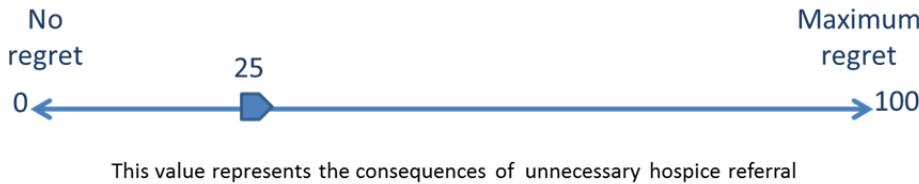
Figure 3. DVAS (Dual Visual Analogue Scales) for the elicitation of threshold probability.

Regret DCA **and extensions**

The clinical problem we face in the situation of hospice referral is how to use reasonably accurate predictions of death, p, coupled with the patient's preferences (as expressed in terms of threshold probability, P_t) to arrive at the optimal decision for a specific individual. The problem is decomposed into three strategies: (1) act based on the prediction model (SUPPORT) (e.g. refer to hospice if $p \geq P_t$ and continue treating otherwise), (2) refer all patients to hospice, and (3) continue current treatment for all patients (i.e. refer no patients to hospice).

Each of these strategies may inflict physiological and/or psychological damages to the patient. Specifically, a patient may suffer harms due to a treatment strategy (e.g. adverse effects) or harms due to the prognostication test (e.g. a test requiring invasive testing). We express these harms as loss in utility associated

17

with actions we may undertake. To that end, we define H_{Rx}, H_{Hosp} and H_{te} as the utility losses due to harms of the treatment, hospice, and prognostic test, respectively.

Figure 4 presents the decision tree describing the overall hospice referral problem. $p = P(D+)$ is the probability that the patient's LE is less than or equal to 6 months as estimated by the prediction model (SUPPORT); $1 - p = P(D-)$ is the probability that the patient's LE is greater than 6 months, and $U_i, i \in [1,4]$, are the utilities corresponding to each of the decision model outcomes (detailed in the previous section). The variables $Hosp$ and Rx correspond to referring a patient to hospice and continuing current curative treatment, respectively. Rg, is the regret associated with an action, e.g. $Rg(Hosp, D-)$ is the regret one may feel if the patient was referred to hospice when his/her LE was greater than 6 months. Finally, te designates that the patient received a prognostication test.

|	Regret	Utilities
Death (p)	$Rg(Hosp,D+)=0$	U_1-H_{hosp}
Refer all — Survival ($1-p$)	$Rg(Hosp, D-)=U_4-U_2-H_{Rx}-H_{hosp}$	U_2-H_{hosp}
Refer: Death $\frac{pTP}{pTP+(1-p)FP}$	$Rg(te, Hosp, D+)=H_{te}$	$U_1-H_{te}-H_{hosp}$
Refer ($pTP+(1-p)FP$): Survival $\frac{(1-p)FP}{pTP+(1-p)FP}$	$Rg(te, Hosp, D-)=U_4-U_2-(H_{Rx}-H_{hosp})+H_{te}$	$U_2-H_{te}-H_{hosp}$
Do not refer: Death $\frac{pFN}{(1-p)TN+pFN}$	$Rg(te, Rx, D+)=U_1-U_3+H_{te}+H_{Rx}-H_{hosp}$	$U_3-H_{te}-H_{Rx}$
Do not refer ($(1-p)TN+pFN$): Survival $\frac{(1-p)TN}{(1-p)TN+pFN}$	$Rg(te, Rx, D-)=H_{te}$	$U_4-H_{te}-H_{Rx}$
Refer none: Death (p)	$Rg(Rx, D+)=U_1-U_3+H_{Rx}-H_{hosp}$	U_3-H_{Rx}
Refer none: Survival ($1-p$)	$Rg(Rx, D-)=0$	U_4-H_{Rx}

Figure 4. Decision tree describing the overall hospice referral process. $p = P(D+)$: probability the patient's LE is less than or equal to 6 months; $1 - p = P(D-)$: probability the patient's LE is greater than 6 months; $U_i, i \in [1,4]$: the utilities corresponding to each of the decision model outcomes; $Hosp$: hospice referral; Rx treatment continuation; Rg : regret associated with an action; H_{Rx} : utility losses due to harms of treatment; H_{Hosp}: utility losses due to harms of hospice; H_{te}: utility losses due to harms of the prognostic test (SUPPORT).

Considering the decision tree in figure 4 we can compute the expected regret associated with each decision in terms of the utilities of each possible outcome as follows (detailed derivation is presented in the Appendix):

$$ERg[Hosp] = (1 - p) * (1 - RRR_{Hosp}) * \frac{P_t}{1-P_t} \tag{3}$$

$$ERg[Rx] = p * (1 - RRR_{Rx}) \tag{4}$$

$$ERg[SUPPORT] =$$
$$\left(1 - RRR_{Hosp} * \frac{\#TP}{n} * RRR_{Hosp} * \frac{\#FP}{n} - RRR_{Rx} * \frac{\#FN}{n} - RRR_{Rx} * \frac{\#TN}{n}\right) * \frac{H_{te}}{U_1-U_3+H_{Rx}-H_{Hosp}} + \left(1 - RRR_{Hosp}\right) *$$
$$\frac{\#FP}{n} * \frac{P_t}{1-P_t} + (1 - RRR_{Rx}) * \frac{\#FN}{n} \tag{5}$$

In addition to harms, equations 3-5 incorporate the effects of treatment and hospice care using measures of *Relative Risk Reduction: RRR_{Rx}* and *RRR_{Hosp}* respectively. The values for these measures are treatment specific can be acquired from the literature. We have incorporated hospice effects because a recent study [11]

has shown that early palliative care for patients with metastatic non-small cell lung cancer could increase survival. The variables *TP*, *FP*, *FN*, *TN* are related to the prognostic capability of the SUPPORT model (see appendix for detailed derivation) [24].

Since the regret of omission and regret of commission have been generalized to include effects and harms related to management strategies and testing, the function of threshold probability (equation 2) becomes:

$$P_t = \frac{1}{1+\frac{U_1-U_3+H_{Rx}-H_{Hosp}}{U_4-U_2-H_{Rx}-H_{Hosp}}} \tag{6}$$

where $U_1 - U_3 + H_{Rx} - H_{Hosp}$ corresponds to the regret associated with not referring the patient to hospice when necessary, and $U_4 - U_2 - H_{Rx} - H_{Hosp}$ corresponds to the regret associated with unnecessary hospice referral.

Choosing the optimal strategy

The optimal strategy is selected as the one which will bring the least amount of regret. The *regret DCA* algorithm expresses the regret associated with each strategy in terms of threshold probability and is implemented as follows [24]:

1. Select a value for threshold probability.
2. Assuming that patients should be treated if $p \geq P_t$ and should not be treated otherwise, compute #TP and #FP for the prediction model.
3. Calculate the $ERg(SUPPORT)$ using equation 5.
4. Calculate $ERg(Rx)$ using equation 4.
5. Compute the $ERg(Hosp)$ using equation 3.
6. Repeat steps 1 – 6 for a range of threshold probabilities.
7. Graph each expected regret function calculated in steps 3-5 against each threshold probability.

At each threshold probability, **the action with the lowest value of expected regret corresponds to the most desired action**. For example, in Figure 5, at a threshold probability equal to 10% (e.g. the patient considers 9 times worse not to be referred to hospice when necessary than to receive an unnecessary hospice referral), the optimal strategy is to refer the patient to hospice.
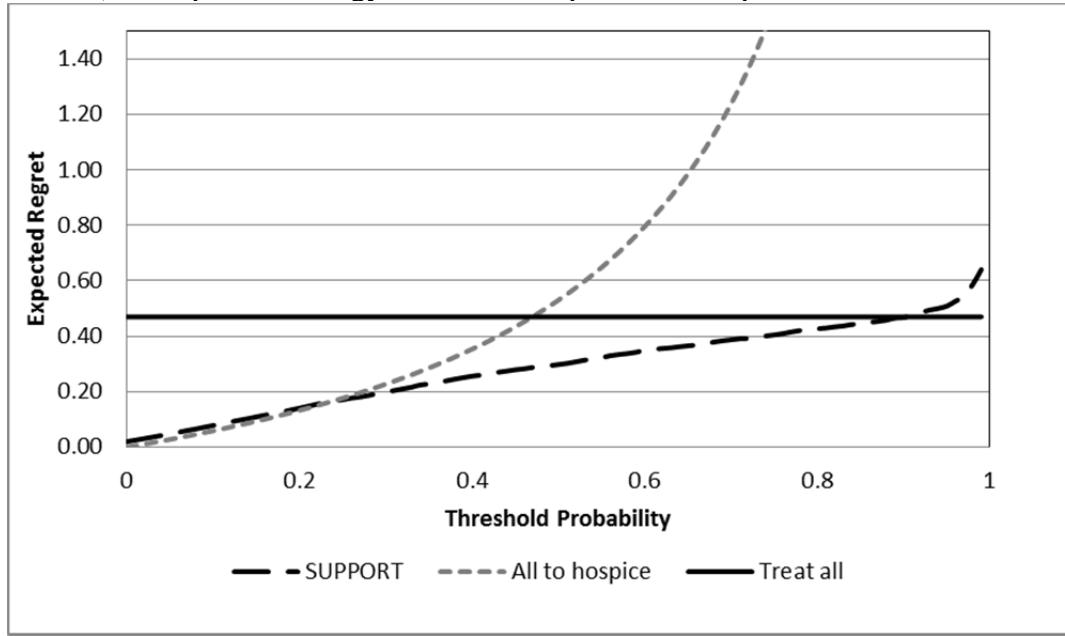

Figure 5. Decision curves for hospice referral. In this figure, $RRR_{Hosp} = 0, RRR_{Rx} = 0, H_{Rx} = H_{Hosp} = H_{te} = 0$. At threshold probability equal to 10%, the optimal decision is refer the patient to hospice; at 40% the optimal decision is to use the SUPPORT model.

Figures 6 and 7 depict 3dimensional representations of the regret associated with alternative decision strategies as they relate to different values of hospice effectiveness (figure 6), treatment effectiveness (figure 7), and harms due to the prognostication test (figure 8). As expected, when the harms due to the prognostication test are increased, then the area of threshold probability at which the prognostication model is the optimal decision is reduced (figure 6). Even though, it is not expected that the SUPPORT model will actually create harms, at least physiological, to the patient, this is not always the case for other diagnostic tests that may be more invasive (e.g. screening for prostate cancer).
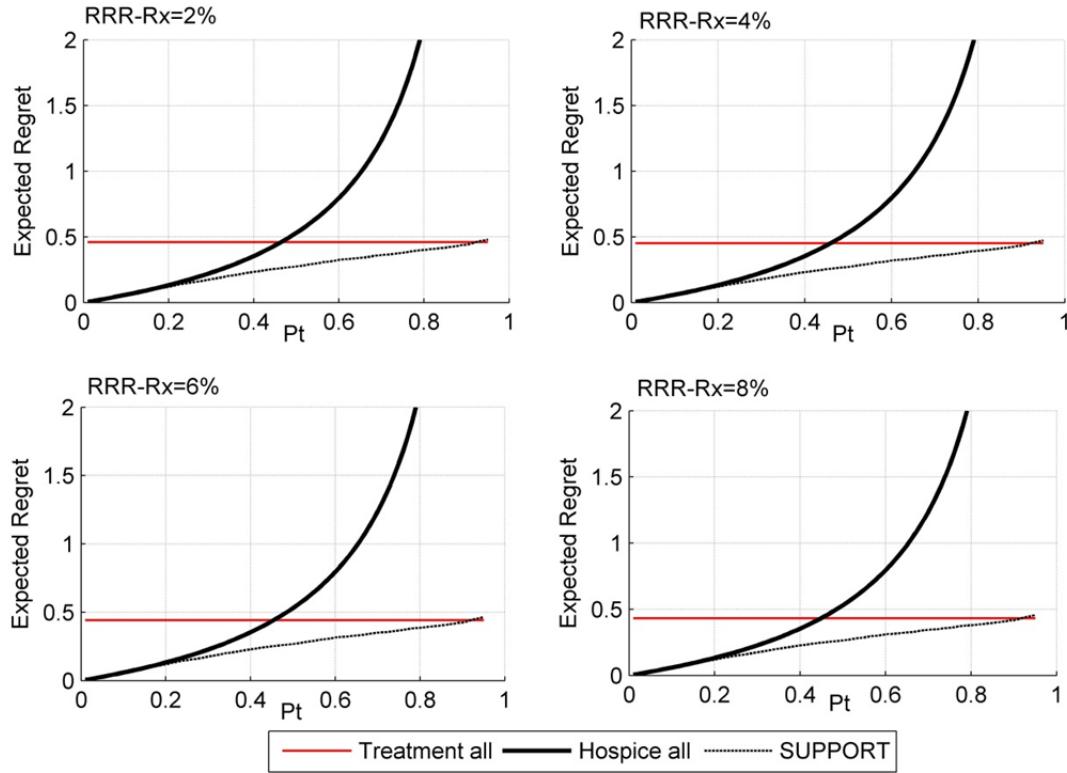
Figure 6. Decision curves as a function of RRR_{RX}. In this figure, $RRR_{Hosp} = 0, RRR_{Rx} = 2$ to $8\%, H_{Rx} = H_{Hosp} = H_{te} = 0$. As the effect of treatment increases, the regret associated with treating all patients and with the SUPPORT model slightly decreases. The strategy of using the SUPPORT model to refer a patient to hospice is the action with the least amount of regret for the wider range of threshold probabilities.
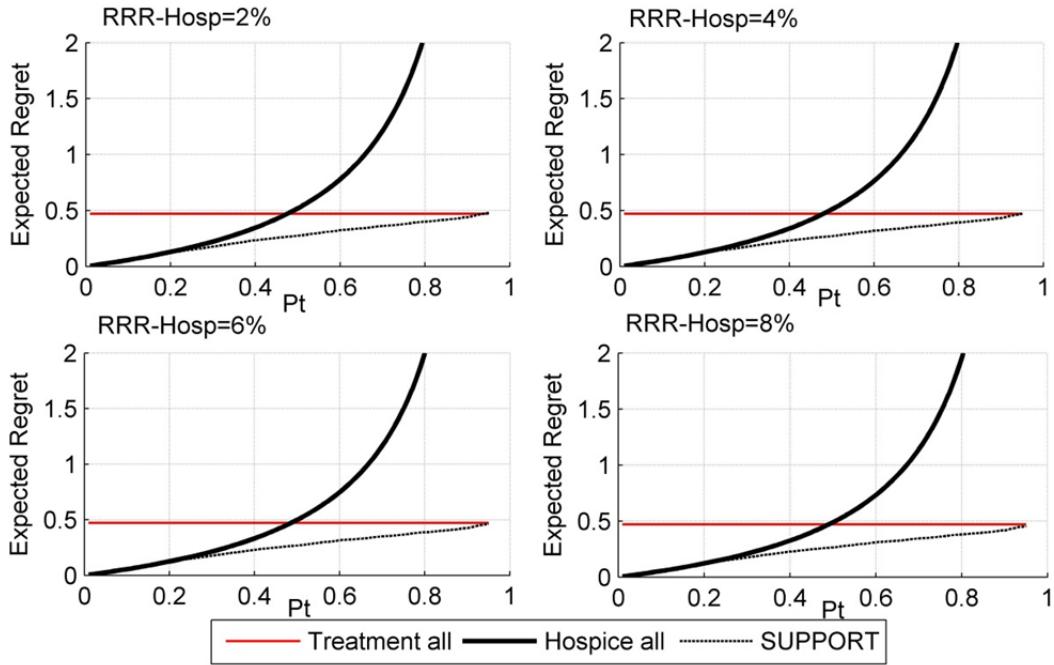
Figure 7. Decision curves as a function of RRR_{Hosp}. In this figure, $RRR_{Hosp} = 2\%$ to 8%, $RRR_{Rx} = 0$, $H_{Rx} = H_{Hosp} = H_{te} = 0$. As the effect of hospice care increases, the regret associated with hospice and with the SUPPORT model slightly decreases. As previously, the strategy of using the SUPPORT model to refer a patient to hospice is the action with the least amount of regret for a wide range of threshold probabilities.
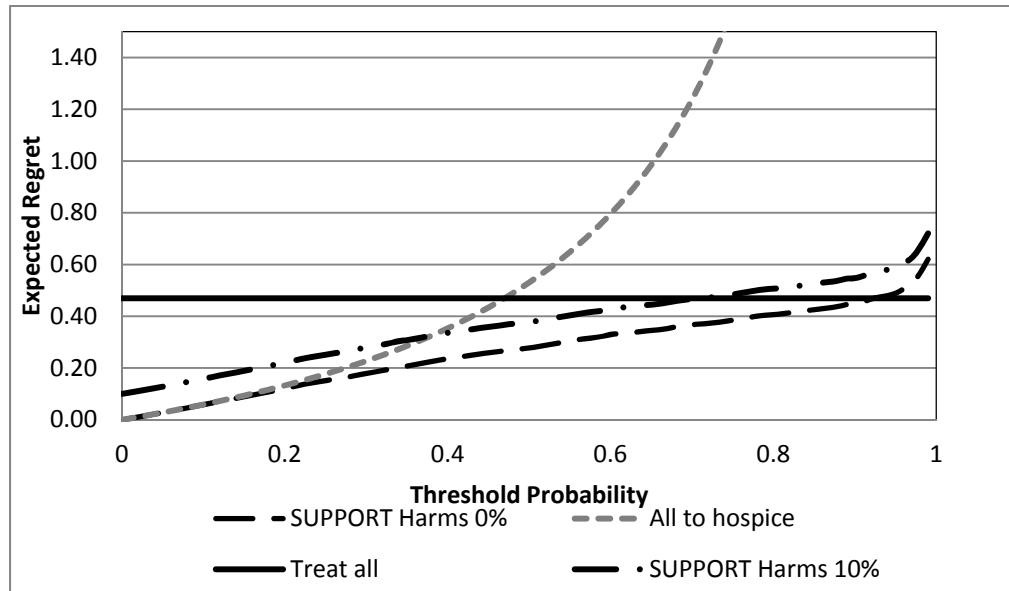
Figure 8. Decision curves as a function of H_{te}. In this figure, $RRR_{Hosp} = 0$, $RRR_{Rx} = 0$, $H_{Rx} = H_{Hosp} = 0$ and $H_{te} = 0\%$ and 10%. As the harms associated to the prediction test increase, so does the expected regret of utilizing the SUPPORT model for hospice referral. Increasing the harms due to treatment or due to hospice care does not have an effect on the decision curves.

As can be seen from figures 5-8, the optimal decision is derived by the SUPPORT model for a rather wide range of threshold probabilities. Therefore, it appears that the SUPPORT model is the superior strategy for the vast majority of decision makers, regardless the effects of the alternative management strategies. However, since the threshold probability expresses the personal preferences of a particular decision maker, it is not

21

unusual for specific patients to have smaller or greater threshold probability values than the majority of decision makers. This is the power of the proposed methodology, which allows for decision making at the individual level. For example, if the decision maker presents a threshold probability greater than ≈92%, the optimal decision would be to continue life-sustaining treatment even if it is deemed not to be effective(figure 5). Similarly, for small values of threshold probability, the desired action would be to refer the patient to hospice.

Decision Support System

As our theoretical discussion highlighted, decisions about life and death are complex and difficult at both the emotional and cognitive level. Therefore, it is not surprising that the SUPPORT model originally failed to improve the quality of care for terminally ill patients despite its reasonable accuracy in prediction of probability of survival in this patient population [17, 20]. Any attempt to focus on a single dimension of the complex hospice referral process is not likely to succeed. An accurate prognostic model is only the first step. Having the apparatus to take into account trade-offs associated with the hospice decision referral while taking into consideration the patients' preferences represent further necessary steps to improve the care of terminally ill patients. In addition, we hypothesize that the SUPPORT intervention failed because it was not available at the point of care in real time. This is because the most desired outcomes are best achieved when decision-making occurs in real-time, at the point of care [33, 34].

To facilitate the decision making process for the hospice referral at bedside, we propose a web-based clinical decision support system (CDSS) that computes the probabilities of survival and death for individual patients using SUPPORT model, elicits personal preferences from patients and/or physicians, and utilizes *regret DCA* to suggest the optimal decision for a particular patient.

Features

Access

Our goal is to develop a CDSS that can be accessed by everyone and from anywhere regardless the operating system one uses. At the same time, it is desirable to develop a system that can eventually be integrated with various healthcare providers' electronic medical records (EMR). We concluded that a web-based implementation would fulfil such requirements.

Data storage

The CDSS performs the required computations without retaining or transmitting sensitive and identifiable information.

Results

In this section we present a prototype of the CDSS, developed to demonstrate the applicability of our theoretical framework for hospice referral. Each subsection describes the results of the methods shown in the previous section in conjunction with the description of the corresponding module. Figure 9 depicts the logical diagram that outlines the operation of the CDSS. Briefly, the operation begins by interviewing the patient or surrogate to elicit his/her threshold probability. Based on the value of threshold probability (equation, the optimal strategy for the particular patient is derived (refer to hospice, continue treatment, or use of prediction model). If the optimal strategy is to follow the prediction model (SUPPORT), then using the equations A1, A2 and 1, the probabilities of survival and death are computed for the particular patient. Finally, the probability of death is contrasted with the patient's threshold probability and the optimal decision is derived (refer to hospice, or continue treatment). At each step described, the patient selects the level of information he/she wishes to be exposed to. For example, the patient may not wish to know his/her threshold probability or probability of death. Instead he/she wishes to know only the optimal decision regarding his/her condition.
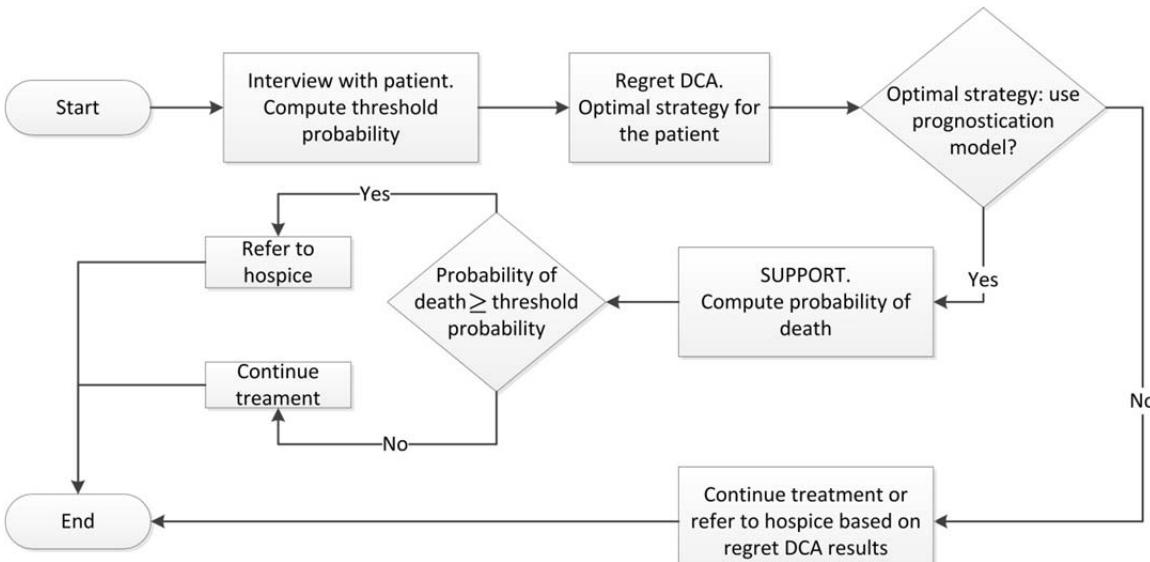
Figure 9. Block diagram outlining the operation of the DSS.

General implementation details

The proposed CDSS is a web-based application residing on the USF Health servers. The web address is http://health.usf.edu/research/ebm/decisionaids.htm. It has been developed based on the Adobe® ColdFusion® application technology and the interface has been designed using html and JavaScript programming languages. The hardware and software requirements from the user's point of view are modest. The system runs on any contemporary computer with net browsing capabilities. However, at this stage the CDSS is not optimized for use with handheld devices. The CDSS consists of 3 different modules as described below.

Elicitation of threshold probability module

The threshold elicitation module consists of the dual visual analogue scales, used to weigh the patient's regret in the case of wrong decisions. Each scale has 100 points where 0 corresponds to no regret and 100 to maximum regret. Depending on the role of the decision maker (e.g. patient/surrogate or physician) two different sets of questions are displayed. These questions are designed to capture the regret of omission and the regret of commission. For the remainder of this paper, we assume that the decision maker is the patient. As in pain scales [35], each visual analogue scale uses facial expressions to graphically represent variations in regret (figure 10). A summary of the decision maker's preferences is presented for final verification. The threshold probability for the particular patient is derived using equation 2, however is not displayed until the decision maker requests it.
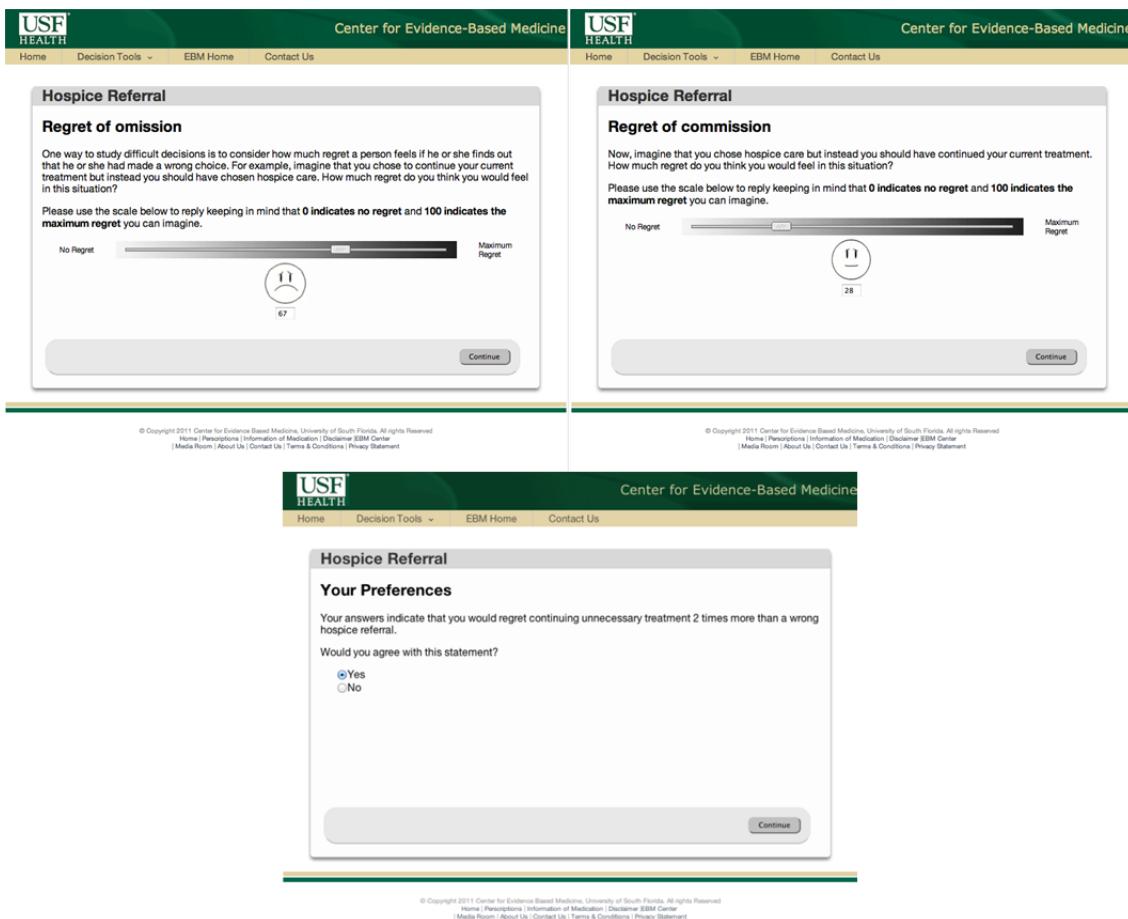
Figure 10. Elicitation of threshold probability. The user (patient/surrogate/physician) weighs the two alternative management strategies in terms of regret.

Decision module

The decision module utilizes the decision maker's threshold probability and the regret DCA methodology to derive the optimal decision. For example, the preferences of the patient depicted in figure 10, correspond to a threshold probability equal to 29%. From figure 5 the strategy that will bring the least amount of regret is to use the prognostication model (SUPPORT) for the hospice referral recommendation. In this case, the decision module initiates the SUPPORT module.

SUPPORT module

If the optimal strategy derived by the decision module is to utilize the prognostication model, the SUPPORT module is enabled. This module (figure 11) is used to compute the probability of death for the particular patient based on the SUPPORT prognostication model. Currently, the user inserts all required information to the CDSS. In the future, this information will be captured automatically from the health care provider's electronic medical records system. Data validation restrictions have been imposed to protect the integrity of the collected data.

Figure 11. SUPPORT user interface. The user enters all information regarding the particular patient to compute the probability of death and survival within the next 6 months. LE results are presented to the patient through the decision justification module after the patient's request.

Once the values of all available variables have been inserted in the corresponding cells, the patient's life expectancy and probabilities of survival and death are computed. The decision module is employed again to display the optimal recommendation.

Decision justification module

The decision justification module explains in detail and at the user's request the reasons that led to a particular recommendation (figure 12). It contains information regarding the decision maker's threshold probability, the optimal strategy associated with the threshold probability and the patient's probability of death (if applicable). Since people often misinterpret probabilities [36], we complement the results presented in terms of probabilities using frequency format (figure 12). The latter format is currently considered the best way to represent favorable and unfavorable facts regarding medical interventions [37]. The justification module is highly technical and should only be reviewed by decision makers who wish to know more about their or their patient's condition.
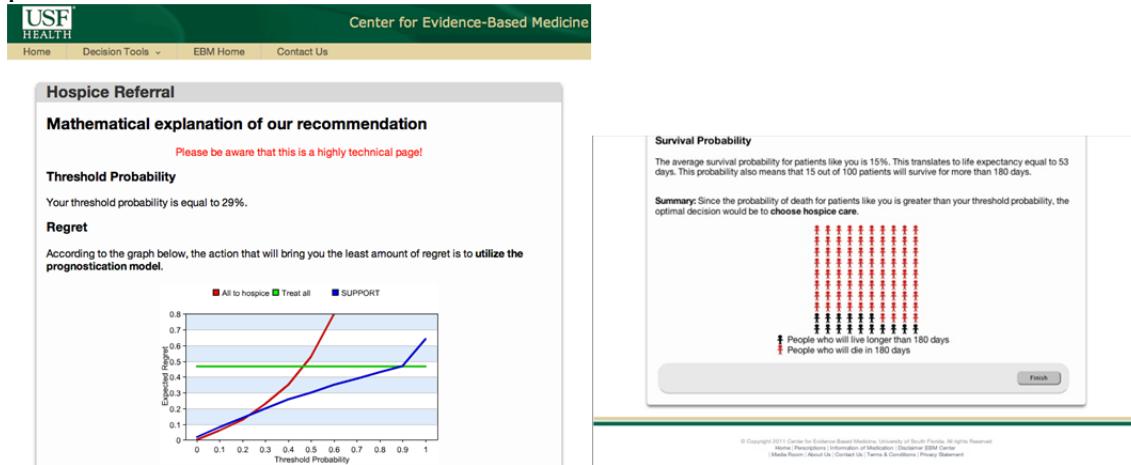
Figure 12. Justification of the hospice referral recommendation. The particular patient depicted has 29% threshold probability at which the optimal strategy is derived by the SUPPORT model. The patient has 85% probability of death in the next 180 days. Therefore, the optimal decision is to be referred to hopsice.

Case Study

Figure 13 summarizes the decision process for a patient whose information is simulated in figures 10-12. The probability of death and threshold probability of this patient have been computed as 29% and 85% respectively. At a 29% threshold probability, the optimal strategy is to use the prediction model for hospice

referral (figure 5). Therefore, since $p > P_t$ the patient should be referred to hospice. For completeness, all possible decision routes are depicted in figure 13. The route corresponding to the specific simulated patient is shown using bold arrows.
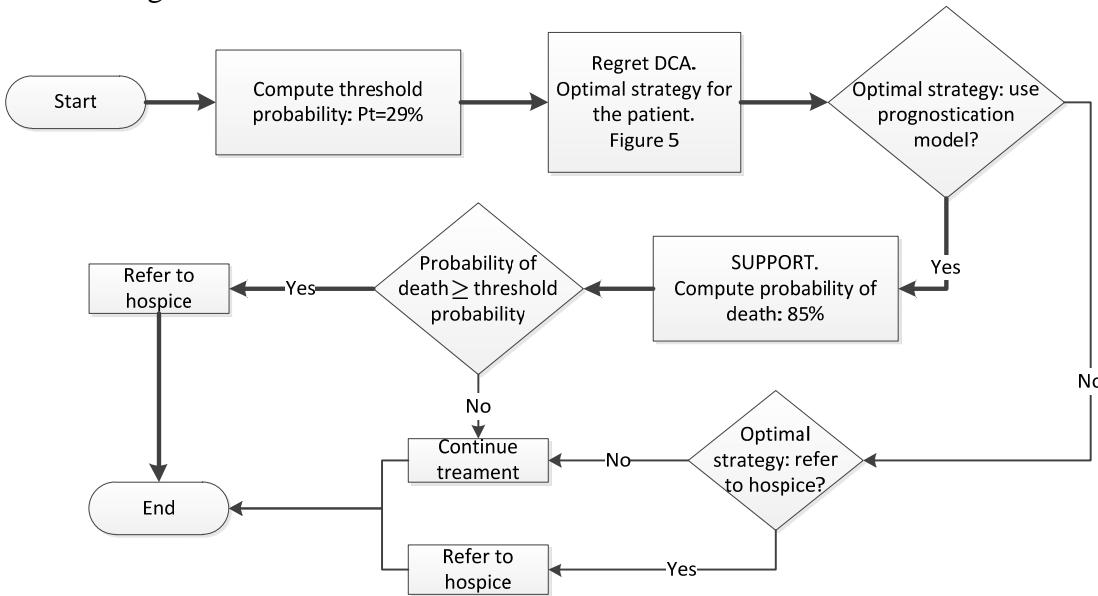

Figure 13. Block diagram summarizing the decision process. The bold route corresponds to the particular simulated patient described in figures 10-12.

Discussion

In this article we describe both the theory and application behind a hospice referral clinical decision support system. To the best of our knowledge, this is the first CDSS that integrates two well established methodologies, one for prognostication (SUPPORT) and the other for decision making (*regret DCA*), to assist with the hospice referral decision-making process.

The recently developed *regret DCA* incorporates the decision maker's preferences towards alternative management strategies from the perspective of regret theory in terms of threshold probability. Such an approach promotes personalized patient care. We anticipate that the regret-based approach is more appropriate for the hospice referral process than other preference elicitation techniques, due to the nature of the problem where there are really no optimal options available- the optimal decision can be only considered as the one with the least regret.

Modern cognitive theories increasingly focus on the so called dual-processing theory in which both intuition (system 1) and analytical, deliberative process (system 2) are important for balancing risks and benefits in the decision-making process [38]. We believe that rational decision-making should take into account both formal principles of rationality and human intuition about good decisions[24, 39, 40]. One way to accomplish this is to use cognitive emotion regret to serve as a link between system 1 and system 2 [24]. By anticipating consequences of our actions and circumstances under which we can live with our mistakes we anticipate that the goal of reconciling formal principles of rationality and human intuitions about good decisions can be met [24, 29, 39, 40]. This is particularly true in the situation of terminally ill patients.

Our web-based CDSS reflects modern cognitive theories to facilitate integration of the decision-making ingredients necessary for hospice referral decisions. The CDSS encapsulates all required information for the hospice referral process into a flexible software that can be used at bedside. Obviously, hospice referral decisions are complex and must be exercised with full compassion and deliberation. We advise against the use of our system as an automatic decision making tool that by-passes important personal interactions between the patient and his/her physician. It is important to stress that the elicitation of the threshold probability as described herein reflects the belief (also captured in recent legislation [41]) that patients and their families want to be told the "truth" about the patients' terminal sickness [22, 41, 42] and that physicians have ethical obligations to

26

share this information with patients and their families [41, 42]. Our system should be understood as an aid to facilitate decisions in terminal phases of patient lives.

Our approach has limitations as well. The main limitation of the proposed system remains the complexity of the SUPPORT model. Currently, the system still requires manual entry of data. In addition, failure to enter all data can jeopardize the accuracy of prediction and therefore, the decision process. To cope with this limitation, we plan to integrate our system into various health providers' EMRs. Based on each EMR, specifically designed queries will be used to retrieve lab values and patient demographics to be fed automatically into our system; a process that will reduce the amount of missing values and input errors.

The second limitation of the proposed system is that empirical data are not available to assess how the system actually works in practice. While we plan to undertake empirical testing of the system described here, we believe that a strong theoretical underpinning will enable better hospice referral decisions even in the current form. This is because our system will essentially operationalize the decision-making process, which is supposed to occur in every day practice. Nevertheless, we need to firstly, identify the system's feasibility in real life settings and ultimately, if it appears to be usable and assessed favourably by all those involved in the hospice-referral decision-making process, to test it in randomized controlled trials against traditional care.

Our future plans include both empirical testing and implementation of multiple additional prognostication models which will be used in parallel to assess optimal decisions regarding hospice referral and take advantage of the DCA methodology. We anticipate that for a different range of threshold probabilities these models may perform better than the SUPPORT model. Furthermore, our intent is to develop a separate version of our CDSS optimized for mobile devices.

Competing interests
The authors have no competing interests to declare.
Authors' contributions
Acknowledgements

References

1. Christakis NA: **Death foretold. Prophecy and prognosis in medical care**: The University of Chicago Press; 2001.
2. Bulkin W, Lukashok H: **Rx for dying: the case for hospice**. *N Engl J Med* 1988, **318**(6):376-378.
3. Christakis NA, Iwashyna TJ: **Spousal illness burden is associated with delayed use of hospice care in terminally ill patients**. *J Palliat Med* 1998, **1**(1):3-10.
4. Goldberg RJ, Mor V, Wiemann M, Greer DS, Hiris J: **Analgesic use in terminal cancer patients: report from the National Hospice Study**. *J Chronic Dis* 1986, **39**(1):37-45.
5. Greer DS, Mor V: **An overview of National Hospice Study findings**. *J Chronic Dis* 1986, **39**(1):5-7.
6. Moinpour CM, Polissar L: **Factors affecting place of death of hospice and non-hospice cancer patients**. *Am J Public Health* 1989, **79**(11):1549-1551.
7. Kane RL, Klein SJ, Bernstein L, Rothernberg R, Wales J: **Hospice role in alleviating the emotional stress of terminal patients and their families**. *Med Care* 1985, **23**(3):189-197.
8. Dawson NJ: **Need satisfaction in terminal care settings**. *Soc Sci Med* 1991, **32**(1):83-87.
9. Kidder D: **The effects of hospice coverage on Medicare expenditures**. *Health Serv Res* 1992, **27**(2):195-217.
10. Mor V, Kidder D: **Cost savings in hospice: final results of the National Hospice Study**. *Health Serv Res* 1985, **20**(4):407-422.
11. Temel JS, Greer JA, Muzikansky A, Gallagher ER, Admane S, Jackson VA, Dahlin CM, Blinderman CD, Jacobsen J, Pirl WF *et al*: **Early palliative care for patients with metastatic non-small-cell lung cancer**. *N Engl J Med* 2010, **363**(8):733-742.

12. Christakis NA: **Timing of referral of terminally ill patients to an outpatient hospice**. *J Gen Intern Med* 1994, **9**(6):314-320.

13. Christakis NA, Escarce JJ: **Survival of Medicare patients after enrolment in hospice programs**. *The New England Journal of Medicine* 1996, **335**(3):172-178.

14. http://www.ssa.gov/OP_Home/ssact/title18/1861.htm

15. Dawes RM, Faust D, Meehl PE: **Clinical versus actuarial judgment**. *Science* 1989, **243**(4899):1668-1674.

16. Hastie R, Dawes RM: **Rational choice in an uncertain world**. London: Sage Publications Inc.; 2001.

17. The Support Investigators: **A Controlled Trial to Improve Care for Seriously Ill Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT)**. *JAMA* 1995, **247**(20):1591-1598.

18. Cook DJ, Guyatt GH, Jaeschke R, Reeve J, Spanier A, King D, Molloy DW, Willan A, Streiner DL, Canadian Critical Care Trials G *et al*: **Determinants in Canadian Health Care Workers of the Decision to Withdraw Life Support From the Critically Ill**. *JAMA* 1995, **273**(9):703-708.

19. Hamel MB, Goldman L, Teno J, Lynn J, Davis RB, Harrell FE, Jr., Connors AF, Jr., Califf R, Kussin P, Bellamy P *et al*: **Identification of Comatose Patients at High Risk for Death or Severe Disability**. *JAMA* 1995, **273**(23):1842-1848.

20. Teno J, Lynn J, Connors A, F, Jr, Wenger N, Phillips RS, Alzola C, Murphy DP, Desbiens N, W.A. K: **The illusion of end-of-life resource savings with advance directives. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment**. *J Am Geriatr Soc* 1997, **45**(4):519-520.

21. Teno JM, Lynn J, Phillips RS, Murphy D, Youngner SJ, Bellamy P, Connors AFJ, Desbiens NA, Fulkerson W, Knaus WA: **Do formal advance directives affect resuscitation decisions and the use of resources for seriously ill patients? SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments.** *J Clin Ethics* 1994, **5**(1):23-30.

22. Knaus WA, Harrell FE, Lynn J, Goldman L, Phillips RS, Connors AF, Dawson NV, Fulkerson WJ, Califf RM, Desbiens N *et al*: **The SUPPORT Prognostic Model: Objective Estimates of Survival for Seriously Ill Hospitalized Adults**. *Annals of Internal Medicine* 1995, **122**(3):191-203.

23. Lynn J, Knaus WA: **Background for SUPPORT**. *J Clin Epidimiol* 1990, **43**(Suppl):1S-4S.

24. Tsalatsanis A, Hozo I, Vickers A, Djulbegovic B: **A regret theory approach to decision curve analysis: a novel method for eliciting decision makers' preferences and decision-making**. *BMC Med Inform Decis Mak* 2010, **10**:51.

25. Vickers A, Elkin E: **Decision curve analysis: a novel method for evaluating prediction models**. *Med Dec Making* 2006, **26**(6):565-574.

26. Djulbegovic B, Hozo I: **When Should Potentially False Research Findings Be Considered Acceptable?** *PLoS Med* 2007, **4**(2):e26.

27. Djulbegovic B, Hozo I, Schwartz A, McMasters KM: **Acceptable regret in medical decision making**. *Med Hypotheses* 1999, **53**(3):253-259.

28. Hozo I, Djulbegovic B: **When is diagnostic testing inappropriate or irrational? Acceptable regret approach**. *Med Dec Making* 2008, **28**(4):540-553.

29. Hozo I, Djulbegovic B: **Will insistence on practicing medicine according to expected utility theory lead to an increase in diagnostic testing?** *Med Dec Making* 2009, **29**:320-322.

30. Bell DE: **Regret in Decision Making under Uncertainty**. *Operations Research* 1982, **30**:961-981.

31. Loomes G, Sugden R: **Regret theory: an alternative theory of rational choice**. *Economic J* 1982, **92**:805-824.

32. Zeelenberg M, Pieters R: **A theory of regret regulation 1.1**. *J Consumer Psychol* 2007, **17**:29-35.

33. Olsen LA, Aisner D, McGinnis MJ: **The learning healthcare system: Workshop summary (IOM roundtable on evidence based medicine)**. Washington: The National Academic Press; 2008.

34. Eden J, Wheatley B, McNeil B, Sox H: **Knowing what works in health care: A roadmap for the Nation**. Washington: The National Academies Press; 2008.

35.   McCaffery M, Beebe A: **Pain: Clinical manual for nursing practice**. Baltimore: V.V. Mosby Company; 1993.
36.   Gigerenzer G, Todd PM, ABC-Research-Group: **Simple heuristics that make us smart**. New York: Oxford University Press; 1999.
37.   Schwartz LM, Woloshin S, Welch HG: **Using a Drug Facts Box to Communicate Drug Benefits and Harms**. *Annals of Internal Medicine* 2009, **150**(8):516-527.
38.   Kahneman D: **Maps of bounded rationality: psychology for behavioral economics**. *American Economic Review* 2003, **93**:1449-1475.
39.   Krantz DH, Kunreuther HC: **Goals and plans in decision making** *Judgement and decision making* 2007, **2**(3):137-168.
40.   Rawls J: **A theory of justice. Revised edition**. Cambridge: Harvard University Press; 1999.
41.   Astrow AB, Popp B: **The Palliative Care Information Act in real life**. *N Engl J Med* 2011, **364**(20):1885-1887.
42.   Mack JW, Weeks JC, Wright AA, Block SD, Prigerson HG: **End-of-life discussions, goal attainment, and distress at the end of life: predictors and outcomes of receipt of care consistent with preferences**. *J Clin Oncol* 2010, **28**(7):1203-1208.

Appendix for this paper

SUPPORT implementation

SUPPORT is implemented in two steps. First, the SUPPORT physiology score is computed based on equation A1.

$$SPS = 259.9\{ARF/MOSF\} + 263.4\{COPD/CHF\}$$

$$+ 241.4\{Cirrhosis/Coma\} + 281.5\{Lung/ColonCancer\}$$

$$- 0.06174\min(PaO_2/FiO_2, 225) - 0.6316\min(MeanBP, 60)$$

$$+ 1.0205WBC - 0.3676(WBC-8)_+ - 0.5631(WBC-11)_+$$

$$+ 0.2691\min(Alb, 4.6) + 0.2312Aresp - 2.362Temp$$

$$+ 1.326(Temp-36.6)_+ + 2.473(Temp-38.3)_+ - 1.579 \times 10^{-1} HR$$

$$+ 9.770 \times 10^{-5}(HR-55)_+^3 - 2.189 \times 10^{-4}(HR-80)_+^3$$

$$+ 1.518 \times 10^{-4}(HR-110)_+^3 - 3.062 \times 10^{-5}(HR-149)_+^3 + 0.9763Bil$$

$$- 0.7481(Bil-7)_+ - 6.8761Cr + 11.6058(Cr-0.600)_+^3$$

$$- 21.8413(Cr-1.000)_+^3 + 10.3574(Cr-1.500)_+^3 - 0.1219(Cr-5.399)_+^3$$

$$- 0.6167096Na + 0.0021118(Na-128)_+^3$$

$$- 0.0036730(Na-135)_+^3 + 0.0006126(Na-139)_+^3 + 0.0009486(Na-148)_+^3$$

$$- 6.278\{COPD/CHF\} \times \min(Alb, 4.6)$$

$$- 11.45\{Lung/ColonCancer\} \times \min(Alb, 4.6)$$

$$+ \{ARF/MOSF\}[-2.3549WBC + 2.7494(WBC-8)_+ - 0.4638(WBC-11)_+]$$
(A1)

where: Alb: albumin; $Aresp$: APACHE III respiration score; Bil: bilirubin; Cr: Creatinine; Na: sodium; PaO_2: partial pressure oxygen in arterial blood; $MeanBP$: mean arterial blood pressure; WBC: white blood cell count in thousands; $Temp$: temperature in Celsius; HR: heart rate per minute; ARF: Acute respiratory failure; $MOSF$: Multiple organ failure; $Cirrhosis$: Cirrhosis; $Coma$: Coma; $Lung$: Lung cancer; $ColonCancer$: Colon cancer; $COPD$: Chronic obstructive pulmonary disease ; CHF: Congestive heart failure. Also:

$$\{disease\ group\} = \begin{cases} 1, & \text{if patient in the disease group} \\ 0, & \text{otherwise} \end{cases}$$

$$(x)_+ = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$WBC = 9$, if $WBC < 9$ and $\{disease\ group\} \neq ARF/MOSF$

$WBC = 40$, if $WBC > 40$

$Cr = 15$, if $Cr > 15$

The second step in implementing the SUPPORT model is to calculate the probability of survival for the individual patient based on equation A2.

$$P\{T \geq t \mid \text{disease group} = i\} = S_i(t)^{e^{X_b'}}$$
(A2)

where T: survival time in days; t: arbitrary time; S described in Table A1 and

$$X_{\hat{b}} = -3.652 + 0.8356\{CHF\} + 0.9257\{Cirrhosis\} + 0.6287\{LungCancer\}$$

$$\pm 1.1803\{MOSFw/Malig\} + 0.01434 Scoma \pm 0.01935 Age + 0.2413 Cancer$$

$$-1.863[Hday + 3.4]^{-1} + 0.08121 SPS + Age[0.015261\{COPD/CHF/Cirrhosis\}$$

$$+0.009047\{Coma\} - 0.008294\{Cancer\}] + Age[-0.012498\{CHF\}$$

$$-0.004578\{Cirrhosis\} - 0.001435\{LungCancer\} - 0.013891\{MOSFw/Malig\}$$

where *Scoma*: SUPPORT coma score (0-100); *MOSFw/Malig*: Multiple organ failure with malignancies; *Hday*: day in hospital when qualified for study; *Cancer*: Cancer by comorbidity or primary disease category (0=no; 1=present; 2= metastatic);

Table A1. Values of Survival (S) as described in equation A2 for different disease types and varying survival times

t	$S_{ARF/MOSF}$	$S_{COPD/CHP/Cirrhosis}$	S_{Coma}	S_{Cancer}
0	0.994	0.998	0.993	0.993
30	0.691	0.889	0.630	0.578
60	0.601	0.837	0.609	0.407
90	0.562	0.800	0.581	0.264
120	0.532	0.772	0.569	0.190
150	0.508	0.751	0.551	0.135
177	0.493	0.733	0.545	0.108

Derivation of the Expected Regret functions

As outlined in the Introduction, seriously and terminally ill patients may reap a number of benefits by the hospice program. Nevertheless, after enrollment into hospice, the patient (or the family, or the physician) may feel that this was a wrong decision, and subsequently may regret it. Similarly, the patient may feel regret for the treatment that he/she continues to receive because it is unnecessary, inappropriate, and/or harmful. Figure 4 represents our hospice decision tree in terms of regret from which we can compute the expected values of regret associated with each strategy as follows:

$$ERg[Hosp] = (1 - p) * (U_4 - U_2 - H_{Rx} - H_{Hosp}) \qquad \text{A3}$$
$$ERg[Rx] = p * (U_1 - U_3 + H_{Rx} - H_{ho}) \qquad \text{A4}$$
$$ERg[SUPPORT] = p * TP * H_{te} + (1 - p) * FP * (U_4 - U_2 - (H_{Rx} - H_{ho}) + H_{te}) + p * FN * (U_1 - U_3 + (H_{Rx} - H_{ho}) + H_{te}) + (1 - p) * TN * H_{te} \qquad \text{A5}$$

The variables *TP*, *FP*, *TN*, *FN* are related to the probabilities $P(p \geq P_t \cap D +)$, $P(p \geq P_t \cap D -)$, $P(p < P_t \cap D -)$ and $P(p < P_t \cap D +)$ respectively, and are estimated as follows:

- $P(p \geq P_t \cap D +) \approx$ the number of patients who will die within 6 months and for whom the prognostic probability is greater than or equal to P_t (with #TP = number of patients with true positive results , $P(p \geq P_t \cap D +) \approx \frac{\#TP}{n}$, where *n* is the total number of patients in the study).
- $P(p \geq P_t \cap D -) \approx$ the number of patients who will survive for longer than 6 months and for whom the prognostic probability is greater than or equal to P_t (with #FP= number of patients with false positive results , $P(p \geq P_t \cap D -) \approx \frac{\#FP}{n}$).
- $P(p < P_t \cap D +) \approx$ the number of patients who will die within 6 months and for whom the prognostic probability is less than P_t (with #FN= number of patients with false negative results, $P(p < P_t \cap D +) \approx \frac{\#FN}{n}$).

- $P(p < P_t \cap D-) \approx$ the number of patients who will survive for longer than 6 months and for whom the prognostic probability is less than P_t (with #TN= number of patients with true negative results, $P(p < P_t \cap D-) \approx \frac{\#TN}{n}$).

To incorporate the effects of alternative treatments (e.g. treatment and hospice care) in equations A3-A5 we use the *Relative Risk Reduction* reported in literature for each strategy as follows:

$$ERg[Hosp] = (1-p) * (1 - RRR_{Hosp}) * (U_4 - U_2 - H_{Rx} - H_{Hosp}) \qquad \text{A6}$$

$$ERg[Rx] = p * (1 - RRR_{Rx}) * (U_1 - U_3 + H_{Rx} - H_{ho}) \qquad \text{A7}$$

$$ERg[SUPPORT] = p * (1 - RRR_{Hosp}) * TP * H_{te} + (1-p) * (1 - RRR_{Hosp}) * FP * (U_4 - U_2 -$$
$$(H_{Rx} - H_{ho}) + H_{te}) + p * (1 - RRR_{Rx}) * FN * (U_1 - U_3 + (H_{Rx} - H_{ho}) + H_{te}) + (1-p) * (1 - RRR_{Rx}) *$$
$$TN * H_{te} \qquad \text{A8}$$

Since $TP + FN = 1$ and $FP + TN = 1$, we have:

$$p * TP + (1-p) * FP + p * FN + (1-p) * TN = p + (1-p) = 1$$

Therefore, equation A8 becomes:

$$ERg[SUPPORT] = \big(1 - p * RRR_{Hosp} * TP - (1-p) * RRR_{Hosp} * FP - p * RRR_{Rx} * FN - (1-p) *$$
$$RRR_{Rx} * TN \big) * H_{te} + (1-p) * (1 - RRR_{ho}) * FP * \Big(U_4 - U_2 - \big(H_{Rx} - H_{Hosp}\big)\Big) + p * (1 - RRR_{Rx}) * FN *$$
$$\big(U_1 - U_3 + (H_{Rx} - H_{Hosp})\big) \qquad \text{A9}$$

Scaling the equations A3, A4 and A9 with the quantity $(U_1 - U_3 + H_{Rx} - H_{ho})$ and replacing the expression $\frac{U_4 - U_2 - (H_{Rx} - H_{ho})}{U_1 - U_3 + H_{Rx} - H_{ho}}$ with $\frac{P_t}{1 - P_t}$, we derive the final equations for the expected regret (equations 3, 4, and 5).

A flexible alternative to the Cox proportional hazard model for prognosticating patient survival

Branko Miladinovic1*, PhD (bmiladin@health.usf.edu) Rahul Mhaskar1, PhD (rmhaskar@health.usf.edu) Sehwan Kim2, PhD (kims@hpchealthcare.org) Ronald Schonwetter2, MD (SchonwetterR@hpchealthcare.org) Benjamin Djulbegovic1,3, MD, PhD

1. Center for Evidence Based Medicine and Health Outcomes Research, University of South

Florida, Tampa, FL, USA

2. HPC Healthcare, Temple Terrace, FL, USA

3. H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA

*Corresponding author

Assistant Professor of Biostatistics

College of Medicine, Center for Evidence-based Medicine

University of South Florida

12901 Bruce B Downs Blvd, MDC27

Tampa, FL 33612

Tel: (813) 396-9614

Fax: (813) 905-8909

Email: bmiladin@health.usf.edu

Number of tables: 3

Number of Figures: 4

Number of references: 38

Word count: 2,774 (Abstract & Text without references)

ABSTRACT

Context. Performance scales are commonly used by clinicians to prognosticate the length of patient survival. The Cox proportional hazards (CPH) model has traditionally been used to perform the external validation of prognostic survival models using patient performance scales. CPH may be suboptimal due to the possible violation of the underlying assumptions and inflexibility to model the baseline survival function.

Objectives. The aim of this study is to externally validate a prognostic model by evaluating the predictive value of the Palliative Performance Scale (PPS) and quantifying its predictive accuracy on a cohort of hospice patients using the Royston-Parmar (RP) family of survival functions. Additionally, the prediction strengths of the RP and CPH models are compared.

Methods. The criteria used to evaluate models' predictive performance were the explained variation statistic R^2, Brier score, scaled Brier score, area under the receiver operating curve, and the Hosmer-Lemeshow goodness of fit test.

Results. A total of 590 patients were studied retrospectively. In addition to PPS, significant predictors of patient survival were age, gender, and cancer status. Calibration and discrimination measures demonstrated that these variables were strongly predictive of patients' survival in the context of the RP family of survival models. The explained variation statistic demonstrated that the RP family of survival functions provided a better fit (R^2 = 0.27; 95% CI: 0.21-0.33) than the CPH model (R^2 = 0.14; 95% CI: 0.10-0.18).

Conclusion. Our results confirmed that PPS, adjusted for age, gender and cancer status is a significant predictor of hospice patient survival. Also, predictive performance and discrimination statistics demonstrated that RP model was as good as, or better than, the CPH model. Researchers are encouraged to consider alternatives to the CPH model in the form of the RP class of flexible parametric functions.

Keywords: Survival prognostication, palliative performance scale, external validation, Royston-Parmar family of survival models.

I. INTRODUCTION

Successful prognostication of patient survival depends on developing and testing prognostic models. Developing a prognostic model entails having accurate patient data for prognosis and selecting clinically relevant candidate predictors and measure(s) of model performance, usually in the context of a multivariable regression survival model1. This process produces patient performance scores that allow for classification of patients into different risk groups. However, the usefulness and validity of a prognostic model is judged by how well it performs for patients that come from different centers2. Validating a prognostic model is generally accepted to mean that given a patient population it works in a data set other than the one it is applied to2,3. In other words, the model needs to be tested using a different data set than the one used to create the model4. It is also generally accepted that the validation process should follow guidelines and that un-validated prognostic models should not be applied in clinical practice4-6. An excellent introduction to the developing and testing clinical prediction models is given by Steyerberg7.

Due to its wide spread use and plausible assumptions, the model of choice for predicting patient survival in a variety of settings (including palliative) has been the Cox proportional hazards (CPH) model8-13. The appeal of the model is its analytic simplicity and that the baseline survival function does not need to be defined apriori -- it is absorbed when the likelihood function is maximized (note that "baseline" refers to zero values of the covariates, not to time equal to zero).

It is possible to estimate the baseline survival function for the CPH model conditional on the estimated regression coefficients. However, this is highly rigid as the smoothing of the underlying function depends on the proportional hazards assumption, which may not be supported by the data. Essentially, the CPH model was designed to measure the effects of covariates on the changing hazard function and not to model patient survival; the usefulness of a prediction model should be measured by how flexible it can be applied to the main model once the external validation is performed. The flexibility of being able to model the (absolute) baseline survival is essential in prognosticating patient survival and a flexible parametric family of functions which allows for parametrically modeling the baseline survival function is more appropriate, especially if the proportional hazards assumption is violated in the CPH14. The baseline survival has for the most part been ignored because it is left undefined in the CPH model. However, it is important in terms of the model's prognostic strength because it specifies the pattern of absolute survival probabilities in the cohort of patients over time.

The goal of this manuscript is to externally validate a hospice patient survival model using a class of flexible Royston-Parmar (RP) parametric functions14 and a well established survival prognostic scale, namely the Palliative Performance Scale (PPS)15. Systematic reviews have shown that the patient palliative performance status is an accurate measure of patient survival in the palliative setting16,17. We intend to show that the RP family of parametric functions allows for a direct and flexible modeling of the baseline survival, and may be formulated so that the impact of the proportional hazard assumption is minimized. Since the CPH model has been widely used in validating prognosticating scales for hospice patient survival9,10,18-22, we will also provide a comparison between RP and CPH models. In the next section we discuss the palliative performance scale and introduce the statistical models.

II. METHODS

Study sample and palliative performance

The patient data were obtained from the Lifepath Hospice and Palliative Care center, licensed since 1983 to serve in Hillsborough County, Florida. Hospice care focuses on pain control and symptom management. We retrospectively extracted data regarding 590 deceased patients starting January 2009 and going backwards. Two research assistants extracted all data necessary to populate the model variables and two faculty members randomly checked 25% of the data for accuracy. The models were tested against observed survival duration.

The Palliative Performance Scale (PPS) was developed and reported by Anderson et al.15 as a measure of palliative patients' functional status. The scale has 11 possible mutually exclusive levels, from PPS of 0% (the patient is dead) to PPS of 100% (the patient is ambulatory and healthy). Numerous studies have studied its

prognostic accuracy of survival in a variety of settings and found it provides meaningful estimates of patient survival10,18,21,23-28 PPS has been found to be both valid and reliable29.

Model selection

When validating a prognostic survival model in the regression framework, most attention has been on the value of the prognostic index based on covariates, while the role of the baseline survival function has been largely ignored. The role of the baseline survival is significant as it quantifies the absolute patient survival probabilities over time.

For a vector of covariates x and parameter vector â, the survival function S(t; x) for the CPH model is commonly expressed as where S0 (t) is the baseline survival function, i.e. survival function when the all the covariates x are equal to zero. In the CPH framework, the estimation of the prognostic index xâ does not require the formulation of the baseline cumulative survival function S0 (t), which itself can be estimated conditional on the covariate estimates. The two popular methods of estimating baseline survival estimators are the Breslow and Kalbfleisch-Prentice estimators30. The two methods give similar results in practice, but can lead to "choppy" estimates of the baseline function and are dependent on the proportional hazards assumption.

When the goal of a survival analysis is to estimate hazard ratios (the effect of covariates on the changing hazard function), the baseline function is of no consequence. The CPH is appropriate as the baseline function gets absorbed when coefficient âs are estimates by the method of partial log likelihood. However, when the goal is to prognosticate patient survival, there is a need for more flexibility in modeling the baseline survival.

An alternative to the CPH is the RP family of models resembles the generalized linear models and can be viewed as a parametric extension Cox proportional hazard models14. The models are framed so they rely on the transformation g(.) where g(.) can be either from the proportional hazard, Aranda-Ordaz or probit families14. Under the proportional hazard link function, the hazard ratio estimates are nearly identical to those estimated under CPH. In essence, the CPH can be viewed as a subset of a more general class of RP functions. The attractive feature of the RP baseline survival function is that its shape is preserved, but its location can vary, which allows for flexible model recalibration. Also, the estimate is implemented on log-time scale is generally gently curved.

In the RP framework, if the proportional hazard assumption is violated, the probit-link function $g(s) = -\ddot{O}^{-1}(s)$ can be applied, where $\ddot{O}^{-1}(.)$ is the inverse standard Normal distribution function. The baseline survival function is approximated and smoothed by a restricted cubic spline function with m (m =0, 1,…, 6) interior knots (see Royston and Parmar14 for details). The optimal number of knots in the RP model can be found using the Akaike Information Criterion (AIC)31. The AIC is defined in the usual manner as - 2Log(likelihood) + 2(No. of model parameters). The methods can be readily implemented in STATA32 statistical software and stpm31 and stpm233 commands.

Predictor variable selection

There is no consensus among researchers as to the best method for selecting predictor variables when developing a prognostic model. We used a combination of criteria to select the best predictors, namely that for a single predictor the selection should be based on 15.7% level of significance34 and that backward elimination should be used to eliminate weak predictors1.

Assessment of model performance

Model performance is the ability of the estimated risk score to predict survival and is assessed using the measures of explained variation, calibration, and discrimination. Calibration refers to how closely the predicted survival at a pre-specified time agrees with the observed survival. The Brier score is a quadratic scoring rule that calculates the differences between the actual outcomes and predicted probabilities3. The Brier score of 0 indicates a perfect model, while 0.25 indicates a non-informative model (the value achieved when issuing a predicted probability of 50% to each patient). The Brier score may be scaled by its maximum to obtain a range between 0% and 100% and have interpretation similar to the Pearson correlation coefficient35. Since calibration is essentially a test of fit, we applied the Hosmer-Lemeshow (HL) test36 on the dead versus alive binary outcome. The HL Chi-square statistic involves grouping of the observations based on the expected probabilities and then testing the hypothesis that the difference between observed and expected events is simultaneously zero for all the groups. This test is equivalent to testing the hypothesis that the observed number of events in each of the groups is equal to the expected number of events based on the fitted model. The higher the HL p-value, the better calibrated the model is.

Discrimination is the ability of the risk score to differentiate between the patients who died versus those who survived at a pre-specified time. The global measure of the model's discriminatory power is the explained variation statistic R2, which measures the variation explained by the fitted model37. Higher values of R2 indicate greater discrimination. A statistic commonly used to summarize discrimination with and without the outcome has been the area under the receiver operating curve (AUC)38, which is a plot of the sensitivity (true positive rate) against 1 – specificity (false positive rate) for consecutive cutoffs of the probability of an outcome. The maximum value of AUC = 1 indicate a perfect prediction model, while a value of AUC = 0.5 indicates that 50% of the patients have been correctly classified (as good as by chance). All statistical calculation were performed using Stata version 11.232

III. RESULTS

The patient characteristic of the retrospective cohort are summarized in Table 1. The cohort consisted of 293 males (49.7%) and 295 females (50.0%), and 2 (0.3%) with unknown gender. All 590 patients were followed until death. The mean, median and range of survival times for the patients by PPS at diagnosis, age, gender, cancer status, and diagnosis category are given in Table 2. The table shows that the median survival was fairly evenly distributed across age groups and gender, but unevenly across the cancer status and initial diagnosis category. There were only 15 total observations for PPS = 60%, 70%, 80%, so they were combined with PPS = 50% to obtain meaningful survival estimates. Fourteen patients had missing values for PPS. The Kaplan-Meier curves stratified by initial PPS level are shown in Figure 1. The curves show good separation indicating that the different risk groups are well defined. The log-rank test for equality of survival curves was highly significant at P = 0.0001. Likewise, when adjacent categories of PPS were compared (PPS 10% vs 20%, 20% vs 30%,..etc),

pair-wise log-rank tests were all significant at P = 0.05 level, except for PPS 40% vs PPS 50% (P =0.394), due to the crossing of two survival curves and longer tail of the PPS 40% group.

The significant predictor variables for the RP model were PPS, age, gender, and cancer status. There was no significant interaction among predictor variables. For both CPH and RP models, we found that in addition to the PPS the other significant predictor variables chosen agreed with previously published results[21]. None of the predictors were continuous and the assumption of linearity was not an issue. The global test based on Schoenfeld residuals indicated that the proportional hazard assumption was violated for both PPS alone (P-value<0.001) and PPS adjusted for age, gender, and cancer status (P-value<0.001), which can also be seen from the unparallel log-plot of curves (Figure 2).

For the RP model, the number of optimal knots was found to be m=1. Due to the violation of the proportional hazards assumption, we used the probit-link function when fitting the RP model. R^2 was higher in the RP model (R^2 = 0.27; 95% CI: 0.21-0.33) than the CPH model (R^2 = 0.14; 95% CI: 0.10-0.18) indicating that the RP model explained significantly more variation in the survival data than CPH. To illustrate the differences for the baseline function, Figure 3 shows plots of the CPH and RP baseline survival functions together with the Kaplan-Meier survival curve. The CPH baseline survival is "choppy" to approximately day 12, while the RP is smooth. The two baseline functions converged at around day 12. The plot of predicted survival (Figure 4) clearly shows better separated scatter bands and better classification of patients into different risk groups under the RP than CPH.

The measures of discrimination and calibration for the two models are given in Table 3. The measures and statistical significance for the differences between the two models are given for days 1, 3, 6, 12, 30, and 60. The measures maintained high predictive ability; however the scaled Brier score increased for RP while it decreased for CPH, indicating a steeper decrease in the maximum predictive probability for CPH and higher calibration of the RP model. This is clinically important if long-term prognostication of survival is desired. Overall, RP was as good as or better than CPH in predicting patient survival.

IV. DISCUSSION

Through external validation of a prognostic survival model and using a flexible Royston-Parmar family of survival models, we showed that the Palliative Performance Scale adjusted for age, gender and cancer status performed well as a predictor of hospice patient survival. This agrees with previously published results[21]. Further, we showed that that the RP family of models may be more appropriate than CPH in modeling survival through its flexible modeling of the baseline survival function, which is especially important as the research moves from externally validating to prognosticating survival. Using the RP flexible baseline function modeling will allow for more precise calibration in the prognostic phase.

There are limitations of our study. The first is that it was confined to model development and the external validation aspects of survival prognostication. The RP model needs to be tested and applied prospectively. The second is that our data contained no censored observations, which is unlikely to occur in most modeling situations. Further work will need to apply the methodology to account for both limitations.

The flexible models discussed could greatly improve the ability of researchers to accurately prognosticate patient survival in the palliative hospice settings. They are well calibrated and their discriminatory power is as good as, or better than, the CPH model. They provide more flexibility to incorporate the baseline survival on the independent data set after external validation is completed.

Reference:

1. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338:b604.

2. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med. Feb 29 2000;19(4):453-473.

3. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. Jan 2010;21(1):128-138.

4. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm? J Clin Oncol. Aug 1 2011;29(22):2951-2952.

5. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.

6. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.

7. Steyerberg EW. Clinical prediction models : a practical approach to development, validation, and updating. New York, NY: Springer; 2009.

8. Burton A, Altman DG. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. Br J Cancer. Jul 5 2004;91(1):4-8.

9. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. BMC Med. 2010;8:21.

10. Downing M, Lau F, Lesperance M, et al. Meta-analysis of survival prediction with Palliative Performance Scale. J Palliat Care. Winter 2007;23(4):245-252; discussion 252-244.

11. Wishart GC, Azzato EM, Greenberg DC, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res. 2010;12(1):R1.

12. Wong J, Taljaard M, Forster AJ, van Walraven C. Does adding risk-trends to survival models improve in-hospital mortality predictions? A cohort study. BMC Health Serv Res. Jul 21 2011;11(1):171.

13. Amdal CD, Jacobsen AB, Tausjo JE, et al. Palliative interventions and prognosis in patients with advanced esophageal cancer. Dis Esophagus. Feb 10 2011.

14. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. Aug 15 2002;21(15):2175-2197.

15. Anderson F, Downing GM, Hill J, Casorso L, Lerch N. Palliative performance scale (PPS): a new tool. J Palliat Care. Spring 1996;12(1):5-11.

16. Chow E, Harth T, Hruby G, Finkelstein J, Wu J, Danjoux C. How accurate are physicians' clinical predictions of survival and the available prognostic tools in estimating survival times in terminally ill cancer patients? A systematic review. Clin Oncol (R Coll Radiol). 2001;13(3):209-218.

17. Lau F, Cloutier-Fisher D, Kuziemsky C, et al. A systematic review of prognostic tools for estimating survival time in palliative care. J Palliat Care. Summer 2007;23(2):93-112.

18. Harrold J, Rickerson E, Carroll JT, et al. Is the palliative performance scale a useful predictor of mortality in a heterogeneous hospice population? J Palliat Med. Jun 2005;8(3):503-509.

19. Olajide O, Hanson L, Usher BM, Qaqish BF, Schwartz R, Bernard S. Validation of the palliative performance scale in the acute tertiary care hospital setting. J Palliat Med. Feb 2007;10(1):111-117.

20. Head B, Ritchie CS, Smoot TM. Prognostication in hospice care: can the palliative performance

scale help? J Palliat Med. Jun 2005;8(3):492-502.

21. Lau F, Downing M, Lesperance M, Karlson N, Kuziemsky C, Yang J. Using the Palliative Performance Scale to provide meaningful survival estimates. J Pain Symptom Manage. Jul 2009;38(1):134-144.

22. Steyerberg EW, Keizer HJ, Habbema JD. Prediction models for the histology of residual masses after chemotherapy for metastatic testicular cancer. ReHiT Study Group. Int J Cancer. Dec 10 1999;83(6):856-859.

23. Fainsinger RL, Demoissac D, Cole J, Mead-Wood K, Lee E. Home versus hospice inpatient care: discharge characteristics of palliative care patients in an acute care hospital. J Palliat Care. Spring 2000;16(1):29-34.

24. Morita T, Tsunoda J, Inoue S, Chihara S. Effects of high dose opioids and sedatives on survival in terminally ill cancer patients. J Pain Symptom Manage. Apr 2001;21(4):282-289.

25. Virik K, Glare P. Validation of the palliative performance scale for inpatients admitted to a palliative care unit in Sydney, Australia. J Pain Symptom Manage. Jun 2002;23(6):455-457.

26. Morita T, Tsunoda J, Inoue S, Chihara S. Validity of the palliative performance scale from a survival perspective. J Pain Symptom Manage. Jul 1999;18(1):2-3.

27. Lau F, Bell H, Dean M, Downing M, Lesperance M. Use of the Palliative Performance Scale in survival prediction for terminally ill patients in Western Newfoundland, Canada. J Palliat Care. Winter 2008;24(4):282-284.

28. Lau F, Maida V, Downing M, Lesperance M, Karlson N, Kuziemsky C. Use of the Palliative Performance Scale (PPS) for end-of-life prognostication in a palliative medicine consultation service. J Pain Symptom Manage. Jun 2009;37(6):965-972.

29. Ho F, Lau F, Downing MG, Lesperance M. A reliability and validity study of the Palliative Performance Scale. BMC Palliat Care. 2008;7:10.

30. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. 2nd ed. Hoboken, N.J.: J. Wiley; 2002.

31. Royston P. Flexible parametric alternatives to the Cox model, and more. Stata Journal. 2001;1(1):1-28.

32. Version 11 [computer program] [computer program]. Version 9. College Station, TX: Stata Corporation; 2010.

33. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. The Stata Journal. 2009;9:265-290.

34. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. Applied Statistics. 1999;48:313-329.

35. Hu B, Palta M, Shao J. Properties of R(2) statistics for logistic regression. Stat Med. Apr 30 2006;25(8):1383-1395.

36. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: Wiley; 2000.

37. Royston P. Explained variation for survival models. The Stata Journal. 2006;6:1-14.

38. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. Apr 1982;143(1):29-36.

Appendix 2

External Validation of Prognostic Models in Terminally Ill Patients

Rahul Mhaskar, MPH, PhD1*, Branko Miladinovic, PhD2*, Athanasios Tsalatsanis, PhD2*,
Alfred Mbah, PhD2*, Ambuj Kumar, MD, MPH3, Kim Sehwan, PhD4*, Ronald Schonwetter,
MD5* and Benjamin Djulbegovic, MD, PhD6

1Center for Evidence-Based Medicine, University of South Florida, Tampa, FL;
2USF, Tampa,FL;
3University of South Florida, College of Medicine, Center for Evidence Based Medicine, Tampa;
4HPC healthcare, Tampa, FL;
5HPC Healthcare, Tampa, FL;
6Center for Evidence-Based Medicine & Health Outcomes Research, University of South Florida, Tampa, FL

Background: Over one million Medicare beneficiaries receive hospice care annually. However, besides the well documented advantages of hospice, many Americans do not enjoy maximum benefit from the hospice care. The fundamental reason for this is related to the inappropriate and poorly timed referral of terminally ill patients to hospice. As a result, many patients die within a few days of referral, while some live many years after the referral was made. Improvement in the accuracy of prognosis translates into superior quality of care. Predictions based on statistical modeling have been shown to be superior to physicians' prognostication. However, very few of these statistical models have been externally validated in terminally ill patients. Here we report the external validation of 5 most commonly used prognostication models in a cohort of terminally ill patients: 1) declining exponential approximation of life expectancy (DEALE) 2) study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT), 3) adjusted palliative performance scale (PPS), 4) adjusted Karnofsky performance scale index (Karnofsky) and 5) adjusted eastern cooperative oncology group performance status (ECOG).

Methods: We retrospectively extracted data from 590 deceased patients enrolled in Tampa Bay Lifepath Hospice and Palliative Care starting January 2009 and going backwards to validate the prognostic models. Two research assistants extracted all data necessary to populate the model variables and two faculty members randomly checked 25% of the data for accuracy. The models were tested against observed survival duration. PPS, Karnofsky and ECOG risk scores were predicted using a flexible family of Royston-Parmar parametric models and adjusted for age, gender and presence of cancer. We utilized several metrics to assess the performance of these models. Specifically, we used the Brier score and scaled Brier score (which is very similar to the Pearson correlation coefficient R^2), the area under the receiver operating characteristic curve (AUROC), and the Hosmer-Lemshow goodness-of-fit p-value (HL).

Results: Brier scores were consistently below the non-informative level of 0.25 and AUROC significantly higher than the non-informative level of 0.5 for the adjusted PPS, Karnofsky and ECOG models (table 1). The HL p-value was consistently greater than 0.1 only for PPS. SUPPORT and DEALE models did not predict fit our data well for survival at day one and month one, two and six. The AUROC takes a value close to 0.5, even though the Brier scores were relatively low and HL p-value greater than 0.05, this value is significantly close to 0.5 for SUPPORT and DEALE models (table 1).

Conclusion: None of the prognostication models accurately predicated survival among our cohort of terminally ill patients. However, PPS consistently performed best in predicting survival in terminally ill patients followed by Karnofsky and ECOG.

	Table 1 Discrimination and model performance statistics in survival prognostication					
	Model performance metrics	**PPS**	**Karnofsky**	**ECOG**	**DEALE**	**SUPPORT**
Day 1	Brier	0.089	0.089	0.106	0.106	0.106
	Brier Scaled	8.80%	4.30%	NA	NA	NA
	AUROC (95%CI)	0.747 (0.68, 0.813)	0.747 (0.693, 0.810)	0.709 (0.648, 0.771)	0.526 (0.467, 0.584)	0.62 (0.56, 0.68)
	HL p-value	0.26	0.17	0.11	0.44	0.58
Day 3	Brier	0.179	0.178	0.293	_	_
	Brier Scaled	16%	15.30%	0%	_	_
	AUROC	0.768 (0.726, 0.810)	0.778 (0.737, 0.818)	0.719 (0.676, 0.761)	_	_
	HL p-value	0.29	0.04	0.57	_	_
Day 6 (Median)	Brier	0.199	0.194	0.363	_	_
	Brier Scaled	20.10%	19.40%	NA	_	_
	AUROC (95%CI)	0.775 (0.739,0.816)	0.787 (0.749, 0.823)	0.721 (0.679, 0.764)	_	_
	HL p-value	0.43	0.008	0.44	_	_
Day 10	Brier	0.179	0.183	0.253	_	_
	Brier Scaled	26.70%	25.90%	4.70%	_	_
	AUROC (95%CI)	0.795 (0.757, 0.834)	0.798 (0.761, 0.836)	0.742 (0.697, 0.786)	_	_
	HL p-value	0.15	0.01	0.61	_	_
Day 30	Brier	0.122	0.127	0.095	0.099	0.1868
	Brier Scaled	37.80%	37.50%	12.30%	2.44%	11.20%
	AUROC (95%CI)	0.781 (0.725,0.838)	0.787 (0.734, 0.839)	0.722 (0.651, 0.794)	0.52 (0.467, 0.573)	0.56 (0.48, 0.64)
	HL p-value	0.62	0.25	0.484	0.92	0.23
Day 60	Brier	0.084	0.088	0.04	0.045	0.2
	Brier Scaled	47.40%	47.70%	18.70%	16.10%	13.80%
	AUROC (95%CI)	0.745 (0.653, 0.837)	0.781 (0.689, 0.871)	0.739 (0.62, 0.858)	0.543 (0.468, 0.616)	0.62 (0.5, 0.74)
	HL p-value	0.29	0.32	0.32	0.32	0.67
Day 180	Brier	0.041	0.05	0.006	0.1	0.27
	Brier Scaled	63.20%	58.60%	31.20%	NA	7.90%
	AUROC (95%CI)	0.55 (0.452, 0.648)	0.51 (0.314, 0.71)	0.59 (0.355, 0.83)	0.7 (0.386, 1)	0.58 (0.04, 1)
	HL p-value	0.59	0.54	0.22	_	0.09

CI: confidence interval, H-L: Hosmer-Lemshow statistics, AUROC: area under the receiver operating characteristic curve